PhD Thesis

Translation and validation of ‘the STarT Back Tool’
– a clinical screening tool for predicting outcome and guiding targeted treatment for patients with low back pain

Lars Morsø, PT, MPH
Research Department, Spine Centre of Southern Denmark,
Hospital Lillebaelt, Middelfart
Institute of Regional Health Services Research, University of Southern Denmark

Faculty of Health Sciences
University of Southern Denmark
2013
### Table of contents

**Preface** ........................................................................................................................................ 1
  Supervisors .................................................................................................................................. 1
  List of publications ....................................................................................................................... 2
  Acknowledgement ........................................................................................................................ 3

**Summary in Danish** .................................................................................................................. 6

**Summary in English** .................................................................................................................. 8

**Framework of the thesis** ............................................................................................................ 10

**Background** ................................................................................................................................. 12

**Objective and aims** .................................................................................................................... 16

**Overall methods** ......................................................................................................................... 17
  Designs ......................................................................................................................................... 17
  Materials ....................................................................................................................................... 18
  Analyses ......................................................................................................................................... 19
  Ethics ............................................................................................................................................ 19

**Study 1:** ‘Translation and discriminative validation of the SBT’ .................................................. 21
  Questions and considerations of the process, part I .................................................................... 24

**Study 2:** ‘The predictive and external validity of the SBT in Danish primary care’ ................. 26
  Questions and considerations of the process, part II ................................................................. 29

**Study 3:** ‘Is the psychosocial profile of people with low back pain seeking care in
Danish primary care different from those in secondary care?’ ............................................... 32
  Questions and considerations of the process, part III ............................................................... 35
Preface

Supervisors

Claus Manniche, Professor, D.M.Sc (main supervisor)
Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt, Middelfart, Institute of Regional Health Services Research, University of Southern Denmark

Hanne B. Albert, MPH, PhD (project supervisor)
Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt, Middelfart, Institute of Regional Health Services Research, University of Southern Denmark

Peter Kent, PhD (project supervisor)
Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt, Middelfart, Institute of Regional Health Services Research, University of Southern Denmark
List of publications

1. Translation and discriminative validation of the STarT Back Screening Tool into Danish.

2. The predictive and external validity of the STarT Back Tool in Danish primary care.
   [Epub ahead of print]

3. Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care?

4. The predictive ability of the STarT Back Screening tool in a Danish secondary care setting.
Acknowledgement

This thesis is based on four cohort studies carried out from 2010 to 2013. The project could not have been undertaken without the political willingness and vision from the Region of Southern Denmark. In 2010, a decision was made to dedicate money to the research of various chronic conditions, including low back pain. Substantial financial support was provided from this fund in the Region of Southern Denmark. Financial support was also generously given by the University of Southern Denmark, the Danish Physiotherapy Association, the Danish Rheumatism Association and the Danish Society for General Practice.

Many people have willingly (and unwillingly) been involved in the process of this project. I would like to express my heartfelt thanks and gratitude to everyone for their support during the completion of my thesis, and I apologise that I cannot thank every one of you by name.

I wish to thank;

Claus Manniche, Professor, D.M.Sc, for his open-minded way of thinking in initiating this project and for his support during the process to develop my academic skills and my physiotherapy career at the Spine Centre.

Hanne B. Albert, Ph.D., for her never-ending optimism, friendship, kind nature and faith in me, for critical and constructive input throughout the last 3 years and for sharing thoughts and ideas in times of difficulty. Thanks for never doubting me.

Peter Kent, Ph.D., for his extensive and highly skilled supervision, for always having time for ‘just one more question’, for having the ability to make me feel competent, even when I was asking the most stupid questions, and for his kindness and support. It has been a privilege to work with you.
Jonathan Hill, Ph.D., for letting me have access to STarT-data and for inviting me to visit Keele University, for his hospitality and kind nature during the stay, for his constructive critical advice, always wanting the best for me.

Lene Ververs, Anne Marie Rosager, Heidi L. Rasmussen and Pernille K. Clausen, the secretaries in the Research Department, for their willingness to help me, no matter how much my requests interfered with their work.

My colleagues in the Research Department, Bodil Arnbak, Rikke K. Jensen, Tue S. Jensen, Per Kjær, Søren O’Niell, Thomas B. Nielsen and Alice Kongsted (the perfect ‘office buddy’) for interest, support and collaboration.

Suzanne Capell, for linguistic editing, revision and constructive feedback.

Lene Marie Isaksen and Dorte Lemvigh from the management of the Spine Centre for their support.

Kirsten Kyvik, head of Institute of Regional Health Services for her sincere interest in the PhD project.

The clinicians at the Spine Centre for their interest through the project and for their patience during my ‘appraisal’ of the SBT.

The Spine Centre ‘home boys’ who always had time for a beer on Friday after work.

John Banke, Flemming Pedersen, the primary care physiotherapy clinics, the GP research unit, Henrik Schroll and DAK-E, for involvement and enthusiasm in collecting the primary care data.

Researchers from the Keele University, Staffordshire, England for kindly providing data from English primary care.
Finally, I would like to thank my friends and family, and most of all the three most important people in my life, my fiancée Camilla and my children Sven and Thea, for their love, support and endless patience.

This thesis is dedicated to my sister Carina.

Lars Morsø, 1 March, 2013
Summary in Danish

Baggrund


Metode

Oversættelsen af SBT blev udført efter metoder anbefalet i internationale guidelines, og den diskriminative validering af SBT blev foretaget, og sammenlignet med tidligere engelske resultater. SBT’s prædiktive værdi i primær sektoren i Danmark blev undersøgt og sammenlignet med resultater fra engelsk primær sektor. Forskelle i den psykosociale profil hos rygpatienter i primær- og sekundærsektor, samt den prædiktive værdi af SBT i sekundær sektoren i Danmark blev udført ved at anvende og sammenligne data fra primær- og sekundær sektor.

Resultater

Den danske oversættelse af SBT var sproglig præcis og kunne anvendes af patienter, til trods for forskelle fundet ved validering af den psykosocial subskala. SBT blev fundet brugbart, med tilstrækkelig diskriminativ evne til at kunne anvendes i primær sektor i Danmark. Den prædiktive
Summary

evne i lav- og mellemrisiko grupperne var i overensstemmelse med fund fra England, hvorimod SBT viste reduceret evne til at forudsige prognose i højrisikogruppen.

Sammenligning af den psykosociale patientprofil hos patienter fra dansk primær- og sekundærsektor viste signifikant højere grad af bevægeangst og katastroferingsadfærd hos patienter fra sekundærsektor, derimod var de mindre ’frygtsomme’ end patienter fra primærsektor. På trods af signifikante forskelle på disse parametre, vurderedes forskellene til at være af en størrelsesorden, som ikke gjorde dem klinisk relevante. Test af den prædiktive evne i sekundær sektor viste, at SBT i mindre grad kunne forudsige prognose ved 6 måneders opfølgning i sekundær sektor sammenlignet med primær sektor.

### Hvad var kendt inden dette Ph.d. projekt?

<table>
<thead>
<tr>
<th>Hvad var kendt inden dette Ph.d. projekt?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• SBT kan identificere modificerbare risiko faktorer i primær sektor.</td>
</tr>
<tr>
<td>• SBT kan klassificere patienter i relevante subgrupper</td>
</tr>
<tr>
<td>• SBT har prognostisk og behandlingsmæssig implikation.</td>
</tr>
<tr>
<td>• Målrettet SBT behandling har vist klinisk og økonomisk effekt.</td>
</tr>
</tbody>
</table>

### Hvad har dette Ph.d. projekt bidraget med?

<table>
<thead>
<tr>
<th>Hvad har dette Ph.d. projekt bidraget med?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Oversættelsen af SBT til dansk er forståelig og anvendelig.</td>
</tr>
<tr>
<td>• Den diskriminative validitet af SBT acceptabel</td>
</tr>
<tr>
<td>• Den prædiktive værdi i primær sektor er acceptabel.</td>
</tr>
<tr>
<td>• Prædiktion af prognose i sekundær sektor reduceret.</td>
</tr>
</tbody>
</table>

**Konklusion**

Samlet set viste resultaterne fra oversættelsen, valideringen og test af den prædiktiv evne, at SBT er et anvendeligt og brugbart klassificeringsredskab i dansk primær sektor. På trods af sammenlignelige psykosociale patient profiler på tvær af sektorer, så var SBT’s evne til at forudsige prognose i sekundær sektor ikke så stærk som i primær sektor.
**Summary in English**

**Introduction**

The STarT Back screening Tool (SBT) is a nine-item patient self-report questionnaire for triage of non-specific low back pain patients in primary care. This short multidimensional questionnaire identifies modifiable risk factors such as pain, activity limitation and psychosocial constructs, and its three-level classification (low, medium, high risk of poor outcome) has prognostic and treatment implications. This project: (i) translated the SBT into Danish, (ii) tested its concurrent validity, (iii) quantified its predictive validity in Danish primary care, (iv) investigated differences in psychosocial characteristics between Danish primary and secondary care settings, and (v) quantified its predictive validity in a Danish secondary care setting.

**Methods**

The translation was performed using methods recommended by international guidelines, and the concurrent validity of the questionnaire was performed cross-culturally using Danish and UK datasets. The predictive validity of the SBT in primary care was described and compared cross-culturally also using data from Danish and UK primary care. Differences in psychosocial characteristics and secondary care predictive validity were studies using data from Danish primary and secondary care.

**Results**

The Danish SBT translation was linguistically accurate and, despite differences found in the performance of the psychosocial sub-scale, the resultant version of the SBT had sufficient patient acceptability and discriminative validity to be used in Denmark. The predictive ability of the low- and

---

<table>
<thead>
<tr>
<th>What was known prior to this PhD project?</th>
</tr>
</thead>
<tbody>
<tr>
<td>- The SBT can identify modifiable risk factors in primary care.</td>
</tr>
<tr>
<td>- The SBT can classify patients into relevant subgroups.</td>
</tr>
<tr>
<td>- The SBT has prognostic and treatment implications.</td>
</tr>
<tr>
<td>- The targeting of treatment has been shown to have positive clinical effectiveness and economic impact.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What does this PhD project add?</th>
</tr>
</thead>
<tbody>
<tr>
<td>- The Danish translation of SBT was linguistically accurate and acceptable to patients.</td>
</tr>
<tr>
<td>- The discriminative validity of SBT was acceptable.</td>
</tr>
<tr>
<td>- The predictive ability in primary care was acceptable.</td>
</tr>
<tr>
<td>- The predictive ability of SBT in secondary care was less.</td>
</tr>
</tbody>
</table>
medium-risk SBT subgroups in Danish primary care was similar to that in UK primary care but was slightly reduced in the high-risk group in DK primary care.

The comparison of patient psychosocial profiles across Danish primary and secondary care settings showed significantly higher movement-related fear and catastrophisation in secondary care but lower anxiety. However, the size of these differences was unlikely to be clinically important. Testing of the SBT predictive validity in secondary care showed it was less able to predict poor outcome at 6-month follow-up in a Danish secondary care setting than in a Danish primary care setting.

**Conclusion**

Collectively, the results from these studies on the translation, discriminative validity and predictive validity of the Danish SBT indicate that it is suitable as a triage tool in primary care. Although there were no clinically important differences in the psychosocial profile of patients between primary and secondary care, the predictive ability of the SBT classification subgroups was weaker in Danish secondary care which there may be many reasons for.
Framework of the thesis

Overall, the PhD project consisted of four studies that each addressed components of the overall objective. From these studies, four papers emerged that describe the creation and validation process of the Danish version of the SBT.

The first component of the objective was to investigate whether the SBT was able to identify subgroups of patients with different risks of poor outcome in Danish primary care. This component was addressed by ‘Translation and discriminative validation of the STarT Back Screening Tool into Danish’ (Paper 1), and ‘The predictive and external validity of the STarT Back Tool in Danish primary care’ (Paper 2).

The second component of the objective was to explore whether the SBT might have some applicability in Danish secondary care, which was addressed by ‘Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care?’ (Paper 3), and ‘The predictive ability of the STarT Back Screening tool in a Danish secondary care setting’ (Paper 4).

The content of this thesis summarises and expands selected background, methods, results and discussion points from the PhD project. Each of the four studies will be summarised and some of the dynamics in the validation process of the Danish SBT will be described. Discussions, questions and considerations that emerged in the process will be addressed in bridging sections of ‘Questions and considerations of the process’, between each of the four studies. This flowchart (Figure 1) will be used to guide the reader through the thesis:
Figure 1. Flowchart of the thesis


**Background**

Low Back Pain (LBP) is a common condition with a lifetime prevalence as high as 80% in adults [1] and a one year prevalence of 55% in a Danish cohort [2]. A 2010 report from the National Institute of Public Health in Denmark based on 173,129 respondents showed that 51.3% had been bothered by back pain within the previous 14 days, including 14.0% who had been ‘very bothered’ [3]. The economic burden of LBP for society is high, and includes the costs of treatment, rehabilitation, days off work, loss of earnings and early retirement. Collectively, this Danish economic burden has been estimated to be EUR 1.73 billion per year, with expenses for treatment alone accounting for EUR .75 billion annually [4]. Besides the economic consequences, LBP also has great consequences for individuals in terms of pain, disability, days off work, social relations, perception of overall health and co-morbidity [3-5].

Over the last few decades, research in the field of LBP has increased [6] with an emphasis on the epidemiology and diagnostics of LBP. However, this effort has been challenged by the heterogeneity of LBP and by difficulty in reaching definitive tissue-specific causes for pain in most individual patients [7-9]. These challenges are reflected in several international guidelines which recommend triaging LBP patients into three broad groups: specific LBP, radiculopathy, and non-specific LBP (NSLBP) [10-16]. In primary care, this method of triage still leaves approximately 80% of patients classified in the group having NSLBP [8], and this group still contains people with highly diverse clinical presentations. This diversity has been shown to influence outcome [17, 18]. In an attempt to address this heterogeneity and to test whether approaches other than the ‘one size fits all’ model result in better patient outcomes, there has been an increased focus in recent years on classification subgroups and various classification models have been suggested [19-21]. Previously, the underlying clinical approach towards NSLBP was found to be focused on structural and...
physical models [22] and it has been argued that, beyond the exclusion of specific serious causes of LBP, no evidence-based agreement exists for the classification and prioritisation of NSLBP patients in primary care. Therefore it has been suggested that alternative methods of classification be considered [23].

Recently, there has been an acceptance of LBP as a multidimensional condition that is highly influenced by psychosocial components [22, 24]. Guidelines also recommend assessment of psychosocial characteristics [16], and several studies have highlighted psychosocial components as being risk factors for poor outcome [12, 24-28]. However, the recognition of psychosocial factors in the daily clinic can be challenging. Firstly, many clinicians feel inadequately trained to assess these characteristics and there is evidence that clinician intuition is not very accurate [29]. Secondly, many clinicians remain uncertain of the impact of these factors on outcome and of how to appropriately manage these factors [30, 31]. Consequently, an increased focus has been on questionnaires validated for detecting psychosocial constructs [26]. Unfortunately, many of these questionnaires are very comprehensive and time-consuming [32]. Therefore, there is renewed interest in questionnaires that screen LBP patients for psychosocial factors in practical ways in daily care settings, especially with the purpose of stratifying prognostic risk in LBP.

The STarT Back Screening Tool (SBT) is a nine-item patient self-report questionnaire for triage of NSLBP patients in primary care [33]. This short multidimensional questionnaire identifies modifiable prognostic factors such as co-morbid pain, activity limitation and several psychosocial constructs - all factors known to be risk factors for the persistence of NSLBP [34]. Answers to the SBT questions create a total-score and sub-score, which classify patients into subgroups of people with low, medium, or high risk of poor prognosis based on symptom complexity. This complexity
Background

is based on the sum of physical and psychosocial characteristics of the individual patient (Appendix 1 – the score-flow) [33]. During the development phase that took place from 2006 to 2008 in the United Kingdom (UK), the SBT underwent a thorough validation process that involved testing its measurement properties, deriving the subgroup cut-points, and describing the predictive and criterion validity of the subscale [35]. This evidence showed that, in the UK, the SBT is a reliable and valid screening tool with adequate discriminative ability to classify patients into relevant subgroups based on poor risk of outcome [35]. Besides being predictive of outcome, the SBT also has treatment implications via subgroup-targeted treatment pathways [36]. The results from a high-quality randomised controlled trial (RCT) in UK primary care showed that subgroup-targeted treatment was more clinically effective and more cost-effective than usual care [37].

In the year 2009, on the basis of the UK validation work and preliminary results from the RCT, the SBT looked appealing as a screening questionnaire for Danish primary care. In parallel with the UK development work, the Region of Southern Denmark had an increasing focus on the assessment of LBP patients in the Region, partly driven by a desire to improve the management of LBP patients in the Region, but also driven by a focus on governmental expenses in the musculoskeletal field. The Region started the development of assessment guidelines for primary care and wanted the SBT to be included in the guidelines to assist GPs in recognising risk of poor outcome and in decision-making about appropriate referral and treatment pathways [38]. Although the SBT had face validity for Danish primary care, at that time-point it had only been validated in UK primary care. No Danish translation of the SBT was available and no definitive results of the effectiveness of targeted treatment in any setting had been published. Many questions about its appropriateness in the Danish context needed to be addressed and there was awareness that the sequence in which these questions were addressed was very important. For example, would a Danish translation of the SBT retain the
discriminative and predictive validity of the original? To ensure an adequate foundation for the Danish version of the SBT, a thorough translation and validation process had to be performed. Parallel to these fundamental considerations about the applicability of the SBT for Danish primary care, there was also a desire to explore the opportunities of expanding the SBT into other care settings, patient populations and time-points of measurement.
Objective

Objective and aims

The overall objective of the thesis was to investigate whether the SBT could identify subgroups of patients and predict risk of poor outcome in Danish primary and secondary care settings.

Therefore, the project had the following four aims:

1. To translate the English version of the SBT into Danish and to test its discriminative validity.

2. To test the predictive and external validity of the Danish version of the SBT in Danish primary care and compare it with the English version of the SBT in UK primary care.

3. To investigate whether the psychosocial profile of patients in Danish primary and secondary care settings were different.

4. To compare the predictive ability of the SBT in a Danish secondary care setting and a Danish primary care setting.
**Overall methods**

**Designs**

Research of prognostic factors aims to identify factors associated with clinical outcomes. This identification might reveal factors that are useful as modifiable targets for intervention [39]. However, often prognostic studies do not reach the high research standards in other research designs [40]. Recently a series of papers proposed four themes as a framework for understanding and improving prognosis research (PROGnosis RESearch Strategy or PROGRESS) [39-42]. This research strategy was proposed to address the gap between the potential of prognostic research and the actual impact, challenges and quality of prognostic research [40]. These challenges and methodological flaws in prognostic modelling have also been previously described in the investigation of LBP and other health conditions [43-45]. Despite these challenges, the potential and opportunities of prognostic models are outlined by the PROGRESS group along with recommendations on how to improve prognosis research [40].

In this context, different layers of prognostic questionnaire validation have been suggested in the literature [45-47]. A fundamental assumption of this PhD project was that its purpose was not to create a new SBT tool in the Danish context but to test the classification validity of a Danish-translated version of the current English language SBT. Therefore, the design of this validation process did not retest the question/factor structure or construct validity of the SBT. Instead, we built on the construct validity work already conducted by Hill et al in the UK [33, 35]. The validation pathway that was chosen in this PhD project was; (i) a cross-cultural validation comparing the discriminative validity of the Danish-translated and original UK versions; (ii) a comparison of the SBT predictive validity for poor outcome at 3 months in Danish and UK primary care cohorts; and
(iii) a comparison of the predictive validity of poor outcome at 6 months in Danish primary and UK secondary care cohorts.

Materials

This PhD project was based in the Medical Department of the Spine Centre of Southern Denmark. Primary care collaborators (the Danish Quality Unit of General Practice, GPs, Physiotherapists and the DAK-E research unit) were involved in the recruitment of patients for three of the four studies. Cohorts from several settings were recruited to broaden the external validation. As the research questions were different in each study, the data used in each study also varied. Two of the studies used cross-sectional data (Studies 1 & 3) and two studies used longitudinal data (Studies 2 & 4). The collection of data in each of the Danish cohorts occurred during the period from March 2010 to October 2012 and some cohorts varied in their outcome time-points. Table 1 gives a visual overview of the Danish and UK cohorts used in each study.

Table 1. Overview of cohorts used for each study

<table>
<thead>
<tr>
<th>Study</th>
<th>Danish translation cohort from secondary care</th>
<th>Danish primary cohort (GP, PT)</th>
<th>Danish secondary predictive cohort</th>
<th>UK primary care development cohort</th>
<th>UK primary care BeBack cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>Baseline only</td>
<td></td>
<td></td>
<td>Baseline only</td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td></td>
<td>Baseline &amp; 3-month outcomes</td>
<td></td>
<td>Baseline &amp; 3-month outcomes</td>
<td></td>
</tr>
<tr>
<td>Study 3</td>
<td>Baseline only</td>
<td>Baseline only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 4</td>
<td></td>
<td>Baseline &amp; 6-month outcomes</td>
<td>Baseline &amp; 6-month outcomes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Patients from Danish primary care, who were 18-65 years old, were recruited at baseline in two ways: (1) From GPs on the basis of relevant diagnostic coding (L03= Back pain, L84= Degenerative changes, L86= Back pain with leg pain), or (2) From physiotherapists using the
criteria suggested in the European NSLBP guidelines [16]. Three and 6-month follow-up data were collected for both primary care settings by way of postal questionnaires. Data from Danish secondary care patients were collected using paper-based baseline and follow-up questionnaires for one study and collected electronically for another. The paper-based questionnaires were consecutively posted until 300 completed baseline questionnaires were returned. Patients recruited electronically were included on the basis of a fully completed baseline SBT. For both the Danish primary and secondary care cohorts, we intentionally did not require restrictive inclusion and exclusion criteria, as we wanted the cohorts to reflect broad clinical practice. The data collection methods and inclusion criteria of the UK primary care cohorts have been described in detail in published studies [33, 48]. More detail on all of the cohorts is described in each of the four papers contained in this thesis.

Analyses

In prognostic research, predictive validity has been tested using a variety of methods [49, 50]. For the studies in this PhD project, we chose to mirror the statistical approaches taken in the original UK development studies, as this allowed comparison of results across cohorts, settings and studies. Dichotomized distribution-based outcome measures used in the UK study, additional relative risk of poor outcome when classified by subgroup and ability of discrimination by using the Area Under the Curve (AUC) statistic was applied in the analyses. Additional regression models were also built for further exploration of predictive differences found between the cohorts.

Ethics

This PhD project was approved by the Scientific Ethics Committee of the Region of Southern Denmark (S-20100036) and all patients gave written informed consent for the use of their data for
Overall methods

research. Permission for collection and storage of data in concordance with the rules by Hospital Lillebaelt was given by the Danish Data Protection Agency (2011-41-6286).
Study 1: ‘Translation and discriminative validation of the SBT’

Aim

The aims of this study were to translate the English version of SBT into Danish and to test its discriminative validity.

Methods

There were two phases in this study: (1) a linguistic and cultural translation phase; and (2) a cross-sectional validation phase of the discriminative ability of the SBT. The first phase was conducted using a convenience sample from secondary care, as we believed it unlikely that the concurrent validity / discriminative ability would be affected by episode duration or care setting. The second phase also included data from the original UK primary care cohort (Table 1a).

Table 1a: Cohorts used for the Danish SBT translation and discriminative validation

<table>
<thead>
<tr>
<th>Danish translation cohort from secondary care</th>
<th>Danish primary cohort (GP, PT)</th>
<th>Danish secondary predictive cohort</th>
<th>UK primary care development cohort</th>
<th>UK primary care BeBack cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline only</td>
<td>Baseline only</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This study followed the translation method recommendations of international guidelines [51-53], which resulted in the following translation process being used (Table 2).

Table 2. Phases in the linguistic and cultural translation

<table>
<thead>
<tr>
<th>Phases</th>
<th>Tasks of the translation process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Liaison with SBT developers</td>
</tr>
<tr>
<td>2</td>
<td>Translation from English to Danish</td>
</tr>
<tr>
<td>3</td>
<td>Back translation from Danish to English</td>
</tr>
<tr>
<td>4</td>
<td>Synthesis</td>
</tr>
<tr>
<td>5</td>
<td>Translation committee consensus?</td>
</tr>
<tr>
<td>6</td>
<td>Pilot testing</td>
</tr>
<tr>
<td>7</td>
<td>Testing of final version</td>
</tr>
</tbody>
</table>
Before initiating the translation process, contact was made with the research team at the Arthritis Research Primary Care Centre at Keele University in Staffordshire, England who developed the SBT. This collaborative link between Keele University and our Danish research group was established to determine whether any other researchers had expressed an intention to undertake the Danish translation work, whether the UK developers were aware of investigations into the appropriateness of the SBT in non-primary care settings, and to request access to their original validation data so that we could undertake comparative studies. They granted us access to data from the original validation sample, which enabled us in this study to compare the cross-sectional concurrent validity across Danish and UK cohorts.

Data that needed to be collected for the Danish cohort consisted of the SBT scores and all the reference standard questionnaires used for the SBT constructs. The reference standards were the Roland Morris Disability Questionnaire (RMDQ) for activity limitation [54, 55], the Tampa Scale for Kinesiophobia (TSK) for fear of movement [56, 57], the Hospital Anxiety and Depression Scale (HADS) for anxiety and depression [58, 59] and the Coping Strategies Questionnaire (CSQ) for catastrophisation [60, 61]. These data allowed us to compare the Danish and the UK cohorts on seven of the nine items included in the SBT. For two items, comparable data were not available - co-morbid pain and bothersomeness - as reference standard questionnaires for these constructs were not readily available. The comparison of the discriminative validity was performed using the Area Under the Curve (AUC) statistic derived from Receiver Operating Curves [62].

**Results**

After a thorough translation process that included minor linguistic adjustments being made during phases one to five of the translation process, the Danish version of the SBT was pilot-tested. During each pilot-test, uncertainty and hesitation were noted by a researcher and the findings were
discussed and adjusted in the plenary group. Pilot-testing was repeated until no further uncertainty was observed. The final Danish version of the SBT was then complete (Appendix 2).

For the concurrent validation process, data from 311 secondary care patients were available. There were minor differences in baseline characteristics across the Danish and the UK cohorts with a higher proportion in the Danish secondary care cohort reporting leg and shoulder/neck pain.

The discriminative ability using AUC was analysed for both cohorts. Overall, the AUC point estimates calculated were similar for five items, but there were differences on three psychosocial sub-score items. Table 3 shows the results for the three sub-score items that differed and the full results are shown in Paper 1.

Table 3. Area under Curve for each SBT question compared with its reference standard
HADS=Hospital Anxiety and Depression Scale, CSQ= Coping Strategy Questionnaire (Full model in Paper 1)

<table>
<thead>
<tr>
<th>Question in SBT</th>
<th>Danish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference Standard Point Estimate (CI95%)</td>
<td>Reference Standard Point Estimate (CI95%)</td>
</tr>
<tr>
<td>6. Worrying thoughts have been going through my mind a lot of the time</td>
<td>HADS ANX .837 (CI95% .792 to .882)</td>
<td>HADS ANX .918 (.894 to .942)</td>
</tr>
<tr>
<td>7. I feel that my back pain is terrible and it’s never going to get any better</td>
<td>CSQ .779 (CI95% .726 to .832)</td>
<td>CSQ .925 (.902 to .948)</td>
</tr>
<tr>
<td>8. In general I have not enjoyed all the things I used to enjoy</td>
<td>HADS DEP .735 (CI95% .678 to .792)</td>
<td>HADS DEP .902 (.876 to .929)</td>
</tr>
</tbody>
</table>

Discussion

The results of the translation and discriminative validation processes showed that the SBT had sufficient patient acceptability and discriminative validity to be used in Danish primary care. Divergence was observed on three psychosocial constructs which could have occurred for a number of reasons: the SBT containing inappropriate screening questions for the Danish context, linguistic inaccuracies, cultural differences, differences in association between screening item and reference
standard, inaccuracies in translation of the reference standard questionnaires, the Danish sample
being from secondary care (as opposed to primary care) or just simple sampling variability across
samples.

**Questions and considerations of the process, Part I**

The translation and discriminative study reassured us of the patient acceptability of the Danish
version and indicated that the SBT had sufficient discriminative validity to be applicable in Danish
clinical practice. The discriminative validity study had tested and confirmed the first component of
the ‘foundation’ of a Danish version of the SBT. However, the study also showed a weaker
discriminative ability in the Danish SBT psychosocial sub-scale, but as there were many potential
sources of that finding, we believed that this should not inhibit us from proceeding with
investigating other aspects of the validity of the Danish SBT.

Although linguistically accurate and discriminatively acceptable, the predictive validity of the SBT
in any Danish care setting had not yet been established. We initially investigated the predictive
validity of the Danish SBT in primary care [47, 63]. For this purpose we needed longitudinal data
from a Danish primary care cohort that was comparable to data from UK primary care. In terms of
the overall SBT validation process, our measuring of its predictive validity in another cohort
(Danish) and at another time-point (the original UK studies used 6-month outcomes but we chose to
study 3-month outcomes), also complied with recommendations that suggest broad validation
criteria should include validation at time-points not previously studied [47, 63]. Creation of a
comparable Danish primary cohort required contact with Danish primary care researchers and the
involvement of GPs and physiotherapy primary care clinics. That collaboration resulted in an
electronic form of the SBT for use in GP practices. This electronic questionnaire was triggered by a
pre-defined diagnostic code when entered into the Danish national medical system by the GP. The
questionnaire was completed during the consultation and a sub-grouping classification was instantly calculated for the GP to use in his/her clinical decision-making. The development of the electronic format and the collection of data occurred within the framework of an audit in general practice in the Region of Southern Denmark. It was not possible to translate the use of this electronic format of the SBT into the physiotherapy setting and so data collection there was by patient self-completion in a paper format.
Study 2: ‘The predictive and external validity of the SBT in Danish primary care’

Aim

The aims of this study were to test the predictive and external validity of the Danish version of the SBT in Danish primary care and compare that with the English version of the SBT in UK primary care.

Methods

Investigation of the predictive ability of the SBT in the context of Danish primary care would clarify whether its ability to predict outcome, based on potentially modifiable prognostic factors, was similar in the UK and Denmark. The predictive validity of the UK SBT had originally been established using 6-month outcomes [33], but 3-month UK data were also available and use of that time-point allowed an opportunity for broader external validation. As 3-month outcomes have been shown to be the most important in the clinical course of LBP in primary care [64, 65] and most Danish primary care patients are seen in that period, this was also a reason for our choosing to investigate the SBT predictive validity for outcomes at that time-point. A Danish primary care cohort consisting of patients from GPs and physiotherapists was recruited. This cohort (n=344) was compared with an existing UK primary care cohort (n=856) from the BeBack study [48] (Table 1b). Descriptive information and standardised questionnaires were extracted from both cohorts at baseline and at 3-month follow-up and entered into a database.

Table 1b. Cohorts used for the predictive validity in primary care

<table>
<thead>
<tr>
<th>Danish translation cohort from secondary care</th>
<th>Danish primary cohort (GP, PT)</th>
<th>Danish secondary predictive cohort</th>
<th>UK primary care development cohort</th>
<th>UK primary care BeBack cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline &amp; 3 month-outcomes</td>
<td></td>
<td></td>
<td>Baseline &amp; 3 month-outcomes</td>
<td></td>
</tr>
</tbody>
</table>
Various methods have been used in testing the predictive validity of patient-reported health questionnaires [49, 50]. We chose to mirror the three statistical methods used in the UK development study [33]. This standardisation allowed us to compare our results with those from previous studies and also facilitates comparison in future studies. Comparison was made between proportions of patients with poor clinical outcome at 3 months stratified by SBT, additional risk of poor outcome by being in a higher risk SBT subgroup was estimated and AUC statistics described the ability to discriminate between people with poor outcome at 3 months on three outcomes.

**Results**

The results from both the Danish and the UK cohorts followed the pattern seen in previous studies [33, 66], with the lowest proportion of patients with poor outcome in the low-risk subgroup and the highest proportion in the high-risk subgroup (Figure 2).

These unadjusted results also indicated that the predictive ability in DK primary care equalled that of the UK for the low- and medium-risk subgroups. However, they also suggested that the predictive ability in the Danish cohort did not have the same magnitude of step increase from medium-risk to high-risk subgroups when compared with the UK cohort. This divergence of results...
Study 2

seemed to be centred round the psychosocial subscale of the SBT and it was our impression that this might be a product of the very different treatment exposure that had occurred between the cohorts.

DAK-E who administered the electronic registry extracted data showing that approximately 60% of the Danish cohort had been exposed to physiotherapy treatment, compared with approximately 18% in the UK cohort. Adjusting for changes in the psychosocial factors over the 3 months in the Danish cohort resulted in the adjusted predictive ability of the high-risk subgroup being almost identical to the unadjusted predictive ability observed in the UK cohort (Table 4). Unfortunately, as change data were not available in the UK data, adjusted results could not be calculated in both cohorts.

### Table 4. The odds of having poor clinical outcome on activity limitation at 3 months by SBT subgroup in the Danish and UK cohorts, estimated using logistic regression. (Full model in Paper 2)

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Danish cohort (n=322)</th>
<th>UK cohort (n=845)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio [CI 95%]</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>Unadjusted model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBT low-risk subgroup</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SBT high-risk subgroup</td>
<td>5.57 [2.97; 10.47]</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>Constant</td>
<td>0.32 [.21; .48]</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Discussion

The results of the predictive study in primary care indicated that the ability to predict increased risk of poor outcome at 3 months in Danish primary care was similar to that seen in UK primary care for the low-risk and medium-risk subgroups, and after adjusting for change in the psychosocial factors,
the predictive ability of the Danish high-risk subgroup was almost identical to unadjusted estimates from the UK cohort. Divergence in predictive ability between the cohorts was centred on the high-risk subgroup, which is based on the psychosocial subscale of the SBT. As seen in Study 1, which had examined the discriminative validity of the SBT, a number of reasons could account for this divergence. Those reasons could include differences in treatment exposure or treatment effectiveness, but could also be due to cultural differences in the influence of psychosocial factors on outcome.

**Questions and considerations of the process, part II**

Overall, the two studies conducted to translate and test the discriminative and predictive validity of the SBT in Danish primary care concluded that the Danish version was a useful prognostic stratification tool. Despite minor differences in discriminative and predictive validity compared with other primary care settings [33, 66, 67], the perception was that SBT had potential to support and guide clinicians in their daily clinical decision-making, although the targeted treatment implications of the SBT subgroups remained untested.

However, the question as to whether the SBT had applicability in other care settings remained unaddressed, although others had speculated on this in the literature [68]. Although investigation of the SBT’s applicability in secondary care was appealing, several considerations were raised within the PhD project group. Firstly, the trajectories of recovery might be different in primary and secondary care in ways beyond that simply attributable to episode duration (Figure 3).
In the secondary care setting of the Spine Centre of Southern Denmark, patients are referred by GPs, chiropractors or medical specialists due to sub-optimal improvement in primary care and patient data indicate that they are more complex and have poorer recovery rates. Given that, we wondered whether screening these patients for poor outcome would also be more complex than in primary care, and whether secondary care screening would require the inclusion of additional or alternative constructs than those contained in the SBT.

The notion of a classifying model based on the prediction of poor outcome makes intuitive sense in a recent-onset episode of LBP, but perhaps it was not as applicable for secondary care patients who were already experiencing persistent pain and who had a higher proportion of leg pain and specific LBP (radiculopathy and central stenosis) [69]. Does it make sense to classify some secondary care patients as being at low risk of poor outcome when they are referred on the basis of experiencing a poor outcome in primary care in the first place?

Secondly, the question emerged as to whether the psychosocial profile of patients differed across these care settings and therefore whether the SBT psychosocial subscale was suitable in secondary care. Prior research suggests that the clinical course of patients is different in primary and secondary care [70] but, although it has been shown that psychosocial factors impact on prognosis and outcome [25, 27, 70], there were very limited data available about whether these psychosocial risk profiles differed across primary and secondary care settings. Similarly, differences in the psychosocial profile of people from primary and secondary care classified by the SBT had not been
previously reported. Therefore, we believed it to be important to investigate potential differences in the psychosocial profile of primary and secondary patients before performing further testing of the predictive ability of the SBT in secondary care.
Study 3: ‘Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care?’

Aim

The aim of this study was to investigate whether the psychosocial profile of patients in Danish primary and secondary care settings was different.

Methods

This cross-sectional study was conducted to determine whether patient profiles on the psychosocial constructs (movement-related fear, catastrophisation, anxiety and depression) included in the SBT were different across primary and secondary care settings. For this study, baseline values from the Danish secondary care cohort in Study 1 and from the Danish primary care cohort in Study 2 were used (Table 1c). Therefore, the study was a secondary analysis of the SBT scores and the full psychosocial construct scores for the five SBT items on the psychosocial subscale.

<table>
<thead>
<tr>
<th>Danish translation cohort from secondary care</th>
<th>Danish primary cohort (GP, PT)</th>
<th>Danish secondary predictive cohort</th>
<th>UK primary care development cohort</th>
<th>UK primary care BeBack cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline only</td>
<td>Baseline only</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A comparison of psychosocial scores was made overall across cohorts and across SBT classification subgroups. Linear regression models were also created for each of the psychosocial constructs to adjust for potentially influential covariates (age, gender, work participation, episode duration, pain intensity and activity limitation).
Results

Although there were no significant differences across care settings in the distribution of patients across the SBT subgroups, a slightly higher proportion of patients were classified in the medium-risk subgroup in primary care (42% vs. 32%). An unexpectedly large proportion of patients from secondary care were classified in the low-risk subgroup (39.2%). Overall, there were significantly higher scores in secondary care on movement-related fear and catastrophisation, but lower scores on anxiety. The observed differences across care settings on the psychosocial constructs were also retained when patients were stratified by SBT subgroup (Figure 4).

* Significant differences

The trend towards increased psychosocial reference standard scores by increased risk SBT classification subgroup that has been reported by others [33] was also found in this study. To further
test for care setting as a dominant variable on these psychosocial constructs, separate linear regression models were made for each construct. When simultaneously entering the independent variables of care setting, age, gender, employment status, pain intensity, activity limitation and duration of episode, the variable of care setting was not significant for any of the psychosocial constructs. When tested in a forward stepwise model, care setting was always entered late in the model. Both these analyses indicate that care setting was not a dominant variable on any of the psychosocial constructs.

**Discussion**

The results from this study indicated small but statistically significant differences on four of five psychosocial constructs across Danish primary and secondary health care settings. Although these differences were also broadly retained when stratified by SBT subgroup, our interpretation was that they were so small in magnitude that they were unlikely to be clinically relevant from a patient perspective. This interpretation was based on previous estimates of minimally important clinical differences [71, 72]. Overall, the trend of increased scores on the psychosocial constructs in higher risk SBT subgroups was similar for both settings and reinforced the construct validity of the SBT. Although the distribution of patients across the three SBT subgroups in primary care was very similar to that reported previously in primary care [33], we noted the surprisingly high proportion of patients allocated to the low-risk subgroup in secondary care. This seems inconsistent with the expectation of these patients recovering well in primary care. There could be several reasons for this high proportion of low-risk patients in secondary care: lack of improvement of low-risk patients in primary care due to inadequate reassurance and information on self-management or over-treatment, the SBT not being able to detect clinical characteristics that are important for the different phase of
LBP in secondary care, and different stages of psychosocial response through the clinical course of LBP.

Questions and considerations of the process, part III

Although some new questions did emerge in this study, the comparison of psychosocial patient profiles across care settings encouraged us to further explore the applicability of the SBT in secondary care. It was reassuring that differences in the psychosocial profiles of patients from these care settings were not large and not all in the same direction. Less clear were the implications for the SBT predictive ability in secondary care, for a large proportion of secondary care patients being classified in the low-risk SBT subgroup.

Therefore, multiple potential aspects influencing the predictive ability of the SBT in secondary care were considered. In Denmark, secondary care is defined as government-funded, specialised care requiring specific referral\(^1\) and we knew from Study 3 that patients from secondary care settings had longer episode duration, more frequent leg pain and greater pain intensity. An earlier study had also shown that there were differences in patient case-mix with an increased proportion of patients at the Spine Centre having specific LBP (radiculopathy and central stenosis) [69]. In addition, we were also thoughtful about any potential influence of the differences in the concurrent validity and predictive ability of the Danish SBT psychosocial subscale noted in Study 1 and Study 2. On the other hand, this would be the first study to contribute knowledge about the predictive ability of the SBT in secondary care and thereby to initiate the first step of testing the SBT in this care setting.

\(^1\) As defined in the Great Danish Encyclopaedia (Published 2005. Gyldendals Forlag)
Study 4: ‘The predictive ability of the SBT in a Danish secondary care setting’

Aim

The aim of this study was to compare the predictive ability of SBT in a Danish secondary care setting with a Danish primary care setting.

Methods

In this study, the secondary care component was conducted using a new cohort (n=960) from the Medical Department of the Spine Centre of Southern Denmark. Patients are referred there for evaluation after sub-optimal improvement in primary care. As 6-month outcomes were believed to be more clinically meaningful for secondary care patients, the secondary and primary care cohorts were designed to contain comparable data at baseline and at 6-month follow-up. The primary care component of the study was a secondary analysis of the physiotherapy subsample (n=172) of the primary care cohort collected for Study 2. Only the physiotherapy subsample was used, as 6-month outcome data were only available for this subsample (Table 1d).

Table 1d. Cohorts used for the predictive validity study in secondary care

<table>
<thead>
<tr>
<th>Danish translation cohort from secondary care</th>
<th>Danish primary cohort (PT only)</th>
<th>Danish secondary predictive cohort</th>
<th>UK primary care development cohort</th>
<th>UK primary care BeBack cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline &amp; 6-month outcomes</td>
<td>Baseline &amp; 6-month outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This study also mirrored the statistical methods used in the original development study conducted in the UK [33], as this allowed us to contextualise the results relative to our previous primary care study and those from the UK. In addition, to explore explanations for the results, we used logistic regression to calculate odds ratios for poor outcome adjusted for baseline differences between the
cohort and also calculated the relative risk of poor outcome using baseline pain intensity and baseline activity limitation as predictors.

**Results**

Overall, there were significant differences at baseline across the two cohorts on duration of episode and pain intensity. In concordance with earlier findings [33, 66], the pattern of increased score on pain intensity and activity limitation across the SBT subgroups from low risk to high risk was also found in both cohorts. At 6-month follow-up, there were differences between the two cohorts with patients in secondary care having higher pain intensity and activity limitation, and these differences between cohorts were retained when stratified by SBT subgroup (Table 5).

Table 5. Outcome at 6-month follow-up for the Danish secondary and primary care cohorts.

<table>
<thead>
<tr>
<th>Secondary care cohort</th>
<th>Total cohort n=960</th>
<th>SBT Low risk n=252 (27.7%)</th>
<th>SBT Medium risk n=296 (32.5%)</th>
<th>SBT High risk n=363 (39.9%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain intensity (0-10 scale)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low back pain, median, (IQR)</td>
<td>4.7 (2-6)</td>
<td>3.0 (2-5)</td>
<td>4.3 (2-6)</td>
<td>5.7 (3-7)</td>
</tr>
<tr>
<td>• Leg pain, median, (IQR)</td>
<td>2.7 (0-5)</td>
<td>1.3 (0-4)</td>
<td>2.7 (0-5)</td>
<td>3.7 (1-6)</td>
</tr>
<tr>
<td><strong>Activity limitation (0-100 scale)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Median, (IQR)</td>
<td>47.8 (28-70)</td>
<td>26.1 (13-48)</td>
<td>47.8 (22-70)</td>
<td>65.2 (35-83)</td>
</tr>
<tr>
<td>• Proportion of patients &gt; 30</td>
<td>656 (69.0%)</td>
<td>120 (47.8%)</td>
<td>209 (71.3%)</td>
<td>292 (81.3%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary care cohort</th>
<th>Total cohort n=144</th>
<th>SBT Low risk n=48 (36.9%)</th>
<th>SBT Medium risk n=52 (40.0%)</th>
<th>SBT High risk n=30 (23.1%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain intensity (0-10 scale)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low back pain, median, (IQR)</td>
<td>3.0 (2-5)</td>
<td>3.0 (2-4)</td>
<td>2.0 (0-4)</td>
<td>4.5 (2-5)</td>
</tr>
<tr>
<td>• Leg pain, median, (IQR)</td>
<td>1.3 (0-4)</td>
<td>0.5 (0-3)</td>
<td>2.0 (1-5)</td>
<td>1.0 (0-5)</td>
</tr>
<tr>
<td><strong>Activity limitation (0-100 scale)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Median, (IQR)</td>
<td>26.0 (9-57)</td>
<td>17.0 (4-26)</td>
<td>30.0 (9-55)</td>
<td>65.0 (25-75)</td>
</tr>
<tr>
<td>• Proportion of patients &gt; 30</td>
<td>51 (40.2%)</td>
<td>9 (20.0%)</td>
<td>21 (46.7%)</td>
<td>18 (69.2%)</td>
</tr>
</tbody>
</table>
The proportion of patients with poor outcome on activity limitation at the 6–month follow-up increased from low risk to high risk in both cohorts. However, a larger proportion in secondary care had poor outcome in all the SBT subgroups, most notably in the low risk, where almost half of the secondary care patients still had an RMDQ score > 30 points (on a 0 to 100 scale). The results also showed that, although statistically significant, the gradient of relative risk in secondary care was not as steep as in primary care (Figure 5), which indicated that the predictive ability of the SBT was weaker in secondary care than in primary care.

As significant baseline differences were found across cohorts on episode duration and pain intensity, estimates of SBT predictive ability were adjusted for these differences. Although in these regression models, episode duration and LBP intensity were significantly associated with outcome, the odds ratios for the SBT subgroups predicting outcome changed only marginally. In contrast, this analysis showed that episode duration was predictive in both cohorts and had an influence that was independent of the predictive ability of the SBT subgroups.

Although it was apparent that the predictive ability of SBT subgroups was less in secondary care than in primary care, we lacked any secondary care reference standards to which the predictive
ability of SBT subgroups could be compared. Therefore, we performed a post hoc analysis using alternative reference standard predictors (categorised baseline pain intensity and categorised baseline activity limitation) and found their predictive ability to be nearly identical to that of the SBT subgroups.

Discussion

The SBT subgroups were less predictive in Danish secondary care than in primary care but were as predictive as similarly categorised baseline scores on pain intensity or activity limitation. The later finding is notable, as baseline pain intensity and activity limitation are known to be strong predictors of outcome [73, 74] and the baseline values of the outcome being investigated (in this case, activity limitation) are usually the strongest predictor [67]. Although there were similar proportions across the cohorts with poor activity limitation at baseline, these proportions were different at 6 months, especially in the low-risk subgroup. This indicates less favourable recovery trajectories in secondary care, as has been shown previously [70]. Based on the predictive ability of the SBT subgroups and the predictive ability of baseline pain intensity and activity limitation, it seems that predicting outcome overall in secondary care is challenging, perhaps due to a combination of the secondary care group’s generally having a less favourable clinical course and perhaps due to greater heterogeneity in the outcomes within the subgroups. In the case of the SBT, it may also be that the identification of increased risk of poor outcome due to psychosocial components is more complex in secondary care. Other explanations could also include differences in treatment exposure or differences in case-mix.
**Overall discussion**

When this PhD project was initiated during 2009-2010, the work describing the development of the SBT in the UK had just been published [33] and only preliminary results from the ‘STarT Back Trial’ [37] existed. Since then, the SBT has gained much attention and at the XII LBP Forum for Research in Primary Care in 2012, the SBT was described as having larger potential in the field of LBP than any other prognostic research for the last 10 years. Since 2009, the SBT has been translated into several languages [75, 76] and tested in a number of studies [66, 67, 77]. Methodological developments in prognostic research have also occurred in this period. Proposals for improved quality criteria for the measurement properties of patient-reported outcomes have been suggested by the COSMIN group [78] [47]. Furthermore, a framework for prognostic research has been recommended by the PROGRESS group [39-42].

The aim of the current project was to investigate whether the SBT was able to identify subgroups of patients predictive of risk of poor outcome in Danish primary and secondary care settings. For that purpose, we chose a project design of performing (i) a cross-cultural validation comparing the discriminative validity of the Danish-translated and original UK versions; (ii) a comparison of the SBT predictive validity for poor outcome at 3 months in Danish and UK primary care cohorts; and (iii) a comparison of the predictive validity of poor outcome at 6 months in Danish primary and UK secondary care cohorts. Prior to this work, an extensive development phase had been completed in the UK [35]. Results from that work indicated that the fundamental stages of developing the SBT had been thoroughly investigated, including the selection of specific modifiable prognostic factors, the testing of measurement properties, the formation of subgroup allocation rules, and the description of the predictive validity of the SBT subscale relative to another psychosocial screening questionnaire [32]. In the context of the PROGRESS framework, the initial UK development work
represents activity in the first three phases of the framework: fundamental prognostic research, prognostic factor research and prognostic model research [40]. Collectively, the validation process of the Danish SBT represents research occurring within the third phase of the PROGRESS framework (prognostic model research) [42].

In this PhD project, we initially built on the construct validity work already conducted by Hill et al in their development studies in the UK [33, 35]. Though it could be theoretically argued that different items might have been appropriate in the Danish SBT, our results from Study 1 did not support that notion and it was not our intention to re-examine the content validity of the SBT. Our strategy of testing the external and predictive validity of an existing prognostic model is strongly concordant with the recommendations of the PROGRESS group [39, 40].

Throughout the process of validity testing, the methods in the original development studies were replicated and the results compared across several Danish and UK cohorts at 3- or 6-month follow-up time-points. This approach allowed us to validate the SBT in different cohorts, in different settings, at different time-points and across national health care systems. This method of validation is recommended [47, 63] and external validation is considered highly important [42]. The choice of the same methods and outcome parameters also creates results more suitable for future systematic review purposes.

The predictive ability of the SBT has been investigated in other studies using alternative statistical approaches [66, 67]. In those studies, continuous measures (SBT sum scores instead of SBT subgroups classification) were used to predict continuous outcome measures (such as RMDQ raw scores rather than distribution-based dichotomised scores). It has been argued that the use of
continuous outcome measures avoids the use of arbitrary cut-off estimates that could lead to misclassification and preserves all the potential information contained within continuous variables [39, 40]. In our opinion, the SBT was designed for the stratification of patients into three subgroups of risk and clinicians use the SBT raw scores only as a means to calculate subgroup membership. As such, the SBT is a predictive prognostic tool, not an explanatory prognostic model [79].

The SBT was explicitly developed as an easy applicable screening tool for daily use in the clinic [35], and its capacity to indicate increased risk has therefore not been reported by regression coefficients or regression line slopes but mostly in the more clinically interpretable terms of relative risk. Therefore, while being fully aware that dichotomised outcomes remove potential information, we believed that modelling the SBT subgroups was more interpretable and relevant for clinicians than modelling risk on continuous scales [43].

This PhD project found that classification into the high-risk subgroup using the Danish SBT was less predictive of poor outcome than in the UK. The unadjusted results in Danish primary care were not as strong as in UK primary care (Study 2) and in Danish secondary care not as strong as in Danish primary care (Study 4). While adjustment for covariates in primary care (Study 2) suggested that care setting (GP/physiotherapy) and change over time in psychosocial factors confounded the unadjusted estimates of SBT predictive ability in Danish primary care, adjustment for selected covariates did not alter the predictive ability in Danish secondary care (Study 4). These results focus upon the psychosocial subscale and question whether the items included in the psychosocial subscale are sufficient in a Danish context. In Study 1, we also examined whether alternative questions from the reference standard questionnaires might have shown stronger concurrent and discriminative validity than those chosen originally for the SBT. We did not find evidence that alternative questions better suited these constructs for Danish patients. It remains
possible, however, that alternative or additional psychosocial constructs might improve the performance of the psychosocial subscale of the Danish version of the SBT. There also may be broader public health and social issues that are not represented in the psychosocial subscale that could be considered as additional prognostic factors in secondary care [64, 80, 81]. For example, in pregnancy-related pelvic pain, it has been shown that socio-demographics are influential on outcome [82]. An over-representation of lower socio-demographics in the secondary care cohort could possibly have influenced the SBT predictive ability in that care setting. So, maybe for the SBT to have better predictive ability in secondary care, broader prognostic factors would need to be included.

Overall, the Danish SBT was not as predictive in secondary care as it was in primary care. However, describing additional risk might not be very useful in a setting where more than 50% of the patients in the reference subgroup (low-risk) do not improve. This reference category leaves little room for the increased risk of poor outcome in the medium- or high-risk subgroups to be relatively large. This was also shown in this setting by the predictive ability for constructs usually considered as strong prognostic factors (categorised baseline activity limitation and pain intensity) to be equivalent to that of the SBT subgroups [73, 74]. This predictive difficulty seems broader than the considerations about the Danish SBT psychosocial subscale, as it applies to all the SBT subgroups.

**Strengths and weaknesses of the PhD project**

The strengths of the project were (i) the collaborative relationship with the developers of the SBT at Keele University in the UK, which ensured a contemporary understanding of the SBT and allowed the project to have access to data from the original UK validation cohorts, (ii) the mirroring of statistical approaches used in the UK studies, as this allowed comparison of results and will
facilitate future synthesis of these results with others, (iii) the use of a translation method recommended by international guidelines, (iv) the use of cross-cultural and cross-care setting comparisons to broaden validity testing, (v) the use of different outcome time-points to broaden external validity testing, and (vi) the use of regression to explore and explain variability in findings.

However, the project also has a number of potential weaknesses. Aspects of this project were conducted in close collaboration with the original developers of the SBT and this potentially could have biased our judgement about the applicability of the SBT in Danish health care. However, the collective oversight of the PhD team, the conducting of the project using internationally recommended methods and the convergence of results across countries are likely to have minimised this potential bias.

A secondary care cohort was used for testing the Danish translation (Study 1) but subsequent results (Study 3) showed there were statistically significant but not clinically important differences in the psychosocial profiles of patients in Danish primary and secondary care settings. It was the view of the project team, that this and other differences between care settings, such as pain intensity and episode duration, were unlikely to impact the concurrent validity / discriminative ability of the SBT in a cross-sectional study design. However, we do not have empirical data to test whether that view is correct.

Some of the project data were collected in paper format and some were collected electronically. The electronic format had not yet been validated and therefore data collected in that format could have contained potential bias. Preliminary results from a new study conducted at the Spine Centre
indicate that SBT data collected in electronic and paper formats are equally valid, but these results are yet to be published.

When collecting data electronically, participating GPs had immediate access to the SBT score and subgroup classification. This could potentially have affected their patient management and thereby have affected treatment exposure and predictive validity. We have data suggesting that at least 60% of the GP patients were referred for physiotherapy and, due to incomplete registrations at the GP practices, this number might have been even higher. However, this electronic SBT scoring is readily available to GPs that opt to use it and we therefore think that the results obtained in Study 2 reflect contemporary Danish primary care.
**Perspectives**

The SBT is a classification tool based on subgroups of increased risk of poor outcome/baseline symptom complexity. The SBT subgroups have prognostic and treatment implications in primary care. This project investigated the discriminative and predictive validity of a Danish-translated version of the SBT. Investigation of the targeted treatment implications of the SBT was beyond the scope of the work undertaken in this PhD. SBT-targeted treatment has been investigated in a large scale, high quality RCT in the UK and was more clinically effective and cost-effective than usual care [37]. Therefore, it would be relevant to investigate the SBT-targeted treatment implications in Danish primary care. Such investigation would also be in concordance with the next phase of the PROGRESS framework that encourages research of stratified medicine [41].

Broader than the SBT approach, previously tested types of targeted treatment have been based on best evidence from the literature and international guidelines [83], but there is minimal evidence that any one type of targeted treatment is more effective than alternatives. Therefore, research that combines other types of targeted treatment with the SBT subgroup classification may be of interest, especially for the high-risk subgroup.

Our data do not support the use of the SBT as a prognostic screening tool in Danish secondary care and suggest that predicting subgroups of poor outcome in this setting is challenging. The complexity of components influencing prognosis might be different from those in primary care and require a more complex screening tool. Investigation of the association between other forms of predictive/prognostic models [20, 84] and the SBT classification might be relevant. In addition, the integration of non-patient self-reported measurement pathways could provide further useful information, such as clinical signs measured by clinicians like neurological signs or movement
patterns and imaging findings (for example selected MRI findings). It may also be the case that predictive models in secondary care need to be specific for different types of case-mix (non-specific pain, radiculopathy, central stenosis). The exploration of more complex predictive models in longitudinal studies might also help in the understanding of trajectories of LBP across care settings and at different time-points - from onset of a LBP episode until the cessation of health care-seeking.

The implementation of the SBT as a guidance tool for decision-making in primary care in the Region of Southern Denmark has been initiated and the SBT has been introduced as a mandatory component of the Region’s treatment guidelines. GPs can choose to use the electronic format of the SBT in their clinics, with automatic subgroup allocation and additional decision guidance during the patient consultation. Further development of the electronic format of SBT is in progress, as is an investigation of any potential impact of using the SBT in different formats (electronic versus paper, self-administered versus clinician-administered).
Overall Conclusions of the PhD project

- The Danish translation of the SBT questionnaire was linguistically accurate.
- The discriminative validity of the Danish SBT was comparable with the English version, though lower discriminative validity was found on three psychosocial questions.
- The SBT had a 3-month predictive ability in Danish primary care that was similar to that in UK primary care (a test of the external validity) but the predictive ability of the high-risk subgroup in Danish primary care was reduced.
- The SBT had sufficient patient acceptability, discriminative and predictive validity to be a suitable prognostic triage tool for LBP patients in Danish primary care.
- There were statistically significant, but probably not clinically important, differences in the psychosocial profile of patients in Danish primary and secondary care settings.
- The SBT was not as strong in predicting outcome at 6 months in a Danish secondary care as compared with a cohort from primary care.
References


(31) Kent PM, Keating JL, Taylor NF. Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. Man Ther 2009 Feb;14(1):88-100.


(35) Hill J. Identifying subgroups among patients with low back pain in primary care: Evaluating the STarT Back Tool. Primary Care and Health Sciences, Keele University; 2008.


(38) Styregruppe i Region Syddanmark. Patientforløbsprogram for Rygområdet i Region Syddanmark. Region Syddanmark; 2010.


Fritz JM, Beneciuk JM, George SZ. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. Phys Ther 2011 May;91(5):722-32.


Manuscript 1:

Translation and discriminative validation of the STarT Back Screening Tool into Danish
Translation and discriminative validation of the STarT Back Screening Tool into Danish

Lars Mörsø · Hanne Albert · Peter Kent · Claus Manniche · Jonathan Hill

Received: 12 January 2011 / Revised: 29 April 2011 / Accepted: 4 July 2011 / Published online: 19 July 2011
© Springer-Verlag 2011

Abstract

Objective The STarT Back Screening Tool (STarT) is a nine-item patient self-report questionnaire that classifies low back pain patients into low, medium or high risk of poor prognosis. When assessed by GPs, these subgroups can be used to triage patients into different evidence-based treatment pathways. The objective of this study was to translate the English version of STarT into Danish (STarT-dk) and test its discriminative validity.

Methods Translation was performed using methods recommended by best practice translation guidelines. Psychometric validation of the discriminative ability was performed using the Area Under the Curve statistic. The Area Under the Curve was calculated for seven of the nine items where reference standards were available and compared with the original English version.

Results The linguistic translation required minor semantic and layout alterations. The response options were changed from “agree/disagree” to “yes/no” for four items. No patients reported item ambiguity using the final version. The Area Under the Curve ranged from 0.735 to 0.855 (CI95% 0.678–0.897) in a Danish cohort (n = 311) and 0.840 to 0.925 (CI95% 0.772–0.948) in the original English cohort (n = 500). On four items, the Area Under the Curve was statistically similar between the two cohorts but lower on three psychosocial sub-score items.

Conclusions The translation was linguistically accurate and the discriminative validity broadly similar, with some differences probably due to differences in severity between the cohorts and the Danish reference standard questionnaires not having been validated. Despite those differences, we believe the results show that the STarT-dk has sufficient patient acceptability and discriminative validity to be used in Denmark.

Keywords STarT Back Screening Tool · Linguistic · Cultural · Translation · Psychometric · Validation

Introduction

Predicting outcome from an episode of low back pain (LBP) is of interest to both the clinical and research communities [1–3]. Generic and specific questionnaires have been developed to inform that process [4–6], but there has been a need for simple tools for clinicians to use in the triage of patients in routine clinical care. The STarT Back Screening Tool (STarT) is a nine-item patient self-report questionnaire, recently validated for triage of non-specific LBP patients in primary care [7]. STarT identifies modifiable prognostic factors from the health domains of pain, activity limitation and psychosocial factors, which are risk factors for persistent non-specific LBP. STarT classifies patients into three groups: low, medium or high risk of poor prognosis, which are based on patients’ symptom
complexity [7]. When assessed by GPs, these groups can be used to assist decisions about appropriate evidence-based treatment pathways. Advice and guidance is recommended for the low-risk group, referral for further treatment that focuses on physical aspects for the medium-risk group and referral for a multidimensional treatment that targets both physical and psychosocial factors for the high-risk group [7].

STarT was developed in Britain and has been translated from English into Norwegian, Dutch, French, Spanish, Welsh, Arabic and Mandarin Chinese [8], but not Danish. The objective of this study was to translate the English version of STarT into Danish and test the discriminative validity of the translated version.

Methods

This study consisted of two stages: (1) a linguistic and cultural translation and (2) a psychometric validation of discriminative ability. They were conducted at the Spine Centre of Southern Denmark, a multi-disciplinary secondary care facility. The method used was based on best practice as recommended by translation guidelines [9–11].

Linguistic and cultural translation

An overview of the phases in the linguistic and cross-cultural translation is shown in Table 1.

Phase 1: liaison with STarT developers

Contact was established with the research group at Keele University that developed STarT. This was to determine whether revisions to the English version were in progress, to request collaboration in the translation project and access to the original validation data. A steering committee was formed consisting of two clinicians/academics whose native language was Danish and two whose native language was English. That committee included a representative of the research group at Keele.

Phase 2: translation (English to Danish)

The original STarT questionnaire was translated from English into Danish by a native Danish-speaking, professional translator. The translator was conceptually introduced to the STarT target audience and health condition and asked to take explanatory notes during the course of the translation process.

Phase 3: back translation (Danish to English)

A back translation of the Danish version was then conducted by an independent native-Danish speaking translator whose qualifications included a university degree in English. The translator was similarly conceptually introduced to the STarT target audience and health condition but the translation occurred without direct content knowledge of the original STarT version. During the translation process, explanatory notes were also taken by the translator.

Phase 4: synthesis

The content of the original and reverse-translated English versions were compared, and differences were noted. These differences were discussed with two independent native English-speaking reviewers, one of whom was a clinician and one who was not. The reviewers commented on the differences and a synthesis of these differences was created.

Phase 5: translation committee

The original English, the Danish and the reverse-translated English versions, plus the synthesis of translation differences were presented to a translation committee, who had been formed to ensure cultural relevance and conceptual equivalence. The committee consisted of seven bilingual people (clinicians/academics and lay people) and included chiropractors, physiotherapists, a surgeon and secretaries. The translation committee discussed differences in translation, whether these reflected linguistic imprecision or cultural differences, and where needed, suggested alternative wording. This process continued until consensus was reached.

<table>
<thead>
<tr>
<th>Phases</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Liaison with STarT developers</td>
<td>• Contact with the developers</td>
</tr>
<tr>
<td>2 Translation</td>
<td>• Formation of steering committee</td>
</tr>
<tr>
<td>3 Back translation</td>
<td>• Translation from English to Danish</td>
</tr>
<tr>
<td>4 Synthesis</td>
<td>• Back translation from Danish to English</td>
</tr>
<tr>
<td>5 Translation committee</td>
<td>• Comparison of translations</td>
</tr>
<tr>
<td>6 Pilot testing</td>
<td>• Review of translated versions</td>
</tr>
<tr>
<td>7 Final version</td>
<td>• Reaching of consensus and development of pilot version</td>
</tr>
<tr>
<td></td>
<td>• Testing in clinical setting</td>
</tr>
<tr>
<td></td>
<td>• Revision of pilot version</td>
</tr>
<tr>
<td></td>
<td>• Testing of final version</td>
</tr>
</tbody>
</table>
Phase 6: pilot testing

A Danish pilot version of STarT was tested with 17 randomly selected LBP patients at the Spine Centre to determine the acceptability and comprehensibility of the translation. The only inclusion criteria were the presence of LBP and sufficient Danish language skills to complete the questionnaire. The pilot testing was conducted by two independent members of the translation committee, both of whom were clinicians. For each patient, their response pattern, hesitation and uncertainty while completing the pilot questionnaire were noted, along with the specific questions involved. Patients were asked about item ambiguity and difficulty. These findings were again discussed in the plenary group and the translation was adjusted, until no further hesitation or uncertainty was observed (until ‘saturation’).

Phase 7: final version

A second wave of testing was conducted on ten randomly selected patients using the revised questionnaire, but no further hesitation or uncertainty was observed or item ambiguity reported. This became our final Danish version of STarT, called the STarT-dk.

Psychometric validation

The discriminative validity of the STarT-dk was described using the Area Under the Curve (AUC) statistic derived from Receiver Operating Curves. The AUC was calculated for the seven items in STarT-dk where Danish versions of reference standard questionnaires were available. Each STarT-dk question was used as the ‘state’ variable (a yes on the screening question being the ‘state’) and was compared with its score on the appropriate reference standard [7]. For comparison, the AUC was also calculated for the English version using the original data used to validate the English version [7].

The reference standards were the Roland Morris Disability Questionnaire [12–14] (activity limitation), Coping Strategies Questionnaire [15, 16] (catastrophising), Tampa Scale for Kinesiophobia [17, 18], (fear of movement) and Hospital Anxiety and Depression Scale [19, 20] (anxiety and depression). All reference standards were administered in Danish. To our knowledge Roland Morris Disability Questionnaire is the only questionnaire which has been translated into Danish using methods that meet current recommendations for cross-cultural adaptation.

Translation of the Tampa Scale for Kinesiophobia into Danish was performed at our clinical facility (forward and backward translation), but the cross-cultural adaptation is incomplete and the work is currently unpublished. The Danish translations of The Hospital Anxiety and Depression Scale were compared and a final Danish version was made with the purpose to approximate the original as much as possible. No further information is given on the translation and validation process [22]. These were the only available Danish language versions of the reference standards used in the English study and it was beyond the scope of the current study to validate the translations available for the reference standards.

The AUC represents the ability of the screening question to discriminate between patients with and without the symptom or sign being assessed. In statistical terms, it is sensitivity (true positive rate) divided by 1-specificity (1-true negative rate) [23]. An AUC of 1.0 is perfect discrimination and an AUC of 0.5 is discrimination no better than chance. AUC was chosen as the statistical approach because STarT is a multidimensional questionnaire consisting of one or two screening questions for each of eight underlying constructs. Therefore, multivariable analysis such as Rasch analysis [24] and other tests of internal validity are not appropriate, as they are designed for unidimensional instruments or instruments with many items measuring each construct.

Data were double-entered independently into a database (Epidata 3.1, http://www.epidata.dk “The EpiData Association” Odense, Denmark) by two research assistants. Missing values were imputed using the multiple imputation feature of PASW 18.0 (formerly SPSS) at default settings. Descriptive statistics (proportions, mean and standard deviation) were used to illustrate cohort characteristics. Differences between cohorts or subgroups were tested using Pearson Chi-square test for binomial or ordinal data and unpaired t test for continuous data. Chi-square tests were performed using Prism 5.0 (Graphpad Software Inc, La Jolla, CA, USA). All other statistical analyses were conducted using PASW 18.0 (IBM Inc., Somers, NY, USA).

Results

Linguistic and cultural translation

During phases 2 and 3, some minor linguistic differences emerged for questions 2 (pain in other body regions), 3 (walking), 4 (dressing) and 7 (catastrophising). These differences were presented to the reviewers in phase 4, who independently believed these differences to be of no consequence. For question 6, ‘Worrying thoughts have been going through my mind a lot of the time’ (English version)
The STarT Back Screening Tool

Patient name: __________________________ Date: ____________

Thinking about the last 2 weeks tick your response to the following questions:

1. My back pain has spread down my leg(s) at some time in the last 2 weeks
   - Disagree 0  Agree 1

2. I have had pain in the shoulder or neck at some time in the last 2 weeks
   - Disagree 0  Agree 1

3. I have only walked short distances because of my back pain
   - Disagree 0  Agree 1

4. In the last 2 weeks, I have dressed more slowly than usual because of back pain
   - Disagree 0  Agree 1

5. It’s not really safe for a person with a condition like mine to be physically active
   - Disagree 0  Agree 1

6. Worrying thoughts have been going through my mind a lot of the time
   - Disagree 0  Agree 1

7. I feel that my back pain is terrible and it’s never going to get any better
   - Disagree 0  Agree 1

8. In general I have not enjoyed all the things I used to enjoy
   - Disagree 0  Agree 1

9. Overall, how bothersome has your back pain been in the last 2 weeks?
   Not at all  Slightly  Moderately  Very much  Extremely
   □ □ □ □ □

Total score (all 9): ___________ Sub Score (Q 5-9): ___________

The translation committee reached consensus on a pilot questionnaire that was subsequently tested in phase 6 on 17 patients. During that testing, patient hesitation and perception of item ambiguity were noted primarily for question 5 and also the response options ‘Disagree’ or ‘Agree’. Several patients argued that some questions could not be answered in Danish as ‘Disagree’/‘Agree’ but rather should be answered as ‘No’/‘Yes’ and some patients suggested there was an important distinction in Danish between ‘Not safe’ and ‘Not wise’. Therefore, the translation committee produced a revised translation that included ‘No/Yes’ response options for four of the questions. The revised questionnaire was tested in phase 7 and as saturation had been achieved, no further linguistic adjustments were made. The English version of the STarT questionnaire is shown in Fig. 1a and the final Danish edition of the STarT is shown in Fig. 1b.

Psychometric validation

During this psychometric validation stage, 513 questionnaires were posted and 311 questionnaires returned—a response rate of 60.6%. The only information available on the 202 non-responders was their age and gender. Non-responders were younger (mean age 46.3 SD 14.5) that responders (mean age 51.4 SD 15.7) (p = 0.003 unpaired t test) and less likely to be women (non-responders 49.5% women, responders 59.5% women, p = 0.026 Chi-square). This resulted in a Danish cohort that was dissimilar to the English cohort in age (6.4 year difference in mean age) but almost identical in gender mix.

The participant characteristics of both the Danish and English cohorts are shown in Table 2. The prevalence of each STarT-dk classification group in the Danish cohort was 39.8% (low), 34.0% (medium) and 26.2% (high). The proportion of patients with leg pain within the last 14 days
was 79.4%, while 31.5% also reported pain in either the neck or shoulders. The mean sum score and standard deviation of the reference standard questionnaires are also reported and show that there were differences between the cohorts on most measures. This is likely to reflect the Danish cohort’s being from the secondary care sector and the English cohort from the primary care sector.

Across the STarT variables in the Danish cohort, there was on average 2.19% missing data (range 5.8–0.6%). Across the reference standard variables this was 2.8% (range 10.3–1.0%). The AUC results calculated with missing values and with imputed values were compared. The largest difference in a point estimate of AUC was 0.08. Therefore, as the results calculated with missing data were almost identical to those using imputed values, only results from the original unimputed data are reported.

Identical analyses were made for seven out of nine STarT questions for both the Danish and the English cohorts. The discriminative ability (AUC) for each of those questions in both cohorts is shown in Table 3. In the Danish cohort, AUC point estimates ranged from 0.735 to 0.855 (CI95% 0.678–0.897) and in the English cohort these point estimates ranged from 0.840 to 0.925 (CI95% 0.772–0.948). Overall, the AUC point estimates calculated for five items were similar between the two cohorts, with overlapping confidence intervals but there were differences on three psychosocial sub-score items. The discriminative ability for pain referral, pain localisation, activity limitation and fear of movement was similar. The divergence was on the anxiety, catastrophisation and depression items, as these screening questions had lower discriminative ability in the Danish cohort than in the English cohort.

Therefore, correlations (Pearson $r$) between each individual question on each of the reference standard questionnaires for anxiety, catastrophisation and depression and the sum score of its questionnaire were calculated. This was to
examine if there was evidence that the STarT question used to screen for that construct in the English version might not be the most appropriate in this cultural setting. The results are shown in Appendix 1.

They indicate that it was uncommon that items other than those in the original STarT version displayed a stronger association with the reference standard and that when it occurred the difference in association was negligible (r < 0.086). How much these would vary across consecutive samples in the same language is not known, but they may be of a similar magnitude. Therefore, there must have been other reasons for this divergence in discriminative ability.

### Discussion

The aim of this study was to translate the English version of STarT into Danish and test the discriminative validity of the translated version. We believe the results show that the STarT-dk has sufficient patient acceptability and discriminative validity to be used in Denmark.

Although the discriminative ability was similar between the two language versions for most items, there was a systematic divergence on three psychosocial items and this may have occurred for a number of reasons. The divergence was on the anxiety, catastrophisation and depression items, as these screening questions had lower discriminative ability in the Danish cohort than in the English cohort. As one explanation for this divergence in discriminative ability might be that the screening questions chosen for the English version of STarT might not be the most appropriate for the Danish version but, as shown in Appendix 1, further analysis did not support that explanation. Another possible reason could be linguistic inaccuracies in the translation. This is not likely, as both the translators and translation committee believe the translation to be linguistically accurate.

As the three items showing divergent discriminative validity were all psychosocial constructs and as the discriminative ability was systematically lower on all three, it raises the question as to whether the divergence was the product of a cultural difference in the way Danish people answer psychosocial questions. This seems unlikely as, compared with the English cohort, the Danish cohort scored higher on some psychosocial constructs and lower on others. As the two cohorts scored differently on these psychosocial constructs, it is also possible that the divergence is due to the association between screening question and reference standard not being linear across the whole scoring range. Another reason could be the presence of inaccuracies in the Danish translation of the reference criteria.

### Table 3  Area under Curve for each STarT question compared with its reference standard

<table>
<thead>
<tr>
<th>Question on STarT</th>
<th>Danish Reference standard point estimate (CI95%)</th>
<th>English Reference standard point estimate (CI95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My back pain has spread down my leg(s) in the last 2 weeks</td>
<td>Single question: ‘localisation of pain’ (leg pain stated yes/no) 0.748 (CI95% 0.692–0.805)</td>
<td>Using a single question on current co-morbid pain sites, positive for leg pain (yes/no) 0.856 (0.784–0.927)</td>
</tr>
<tr>
<td>2. I have had pain in the shoulder or neck at some time in the last 2 weeks</td>
<td>Sum score of pain localisation 0.793 (CI95% 0.742–0.845)</td>
<td>Pain sites* 0.898 (0.842–0.955)</td>
</tr>
<tr>
<td>3. I have only walked short distances because of my back pain</td>
<td>RMDQ 0.846 (CI95% 0.804–0.889)</td>
<td>RMDQ 0.880 (0.821–0.938)</td>
</tr>
<tr>
<td>4. In the last 2 weeks, i have dressed more slowly than usual because of back pain</td>
<td>RMDQ 0.855 (CI95% 0.814–0.897)</td>
<td>RMDQ 0.846 (0.772–0.920)</td>
</tr>
<tr>
<td>5. It’s not really safe for a person with a condition like mine to be physically active</td>
<td>TSK 0.775 (CI95% 0.714–0.837)</td>
<td>TSK 0.840 (0.770–0.908)</td>
</tr>
<tr>
<td>6. Worrying thoughts have been going through my mind a lot of the time</td>
<td>HADS ANX 0.837 (CI95% 0.792–0.882)</td>
<td>HADS ANX 0.918 (0.894–0.942)</td>
</tr>
<tr>
<td>7. I feel that my back pain is terrible and it’s never going to get any better</td>
<td>CSQ 0.779 (CI95% 0.726–0.832)</td>
<td>CSQ 0.925 (0.902–0.948)</td>
</tr>
<tr>
<td>8. In general I have not enjoyed all the things I used to enjoy</td>
<td>HADS DEP 0.735 (CI95% 0.678–0.792)</td>
<td>HADS DEP 0.902 (0.876–0.929)</td>
</tr>
<tr>
<td>9. Overall, how bothersome has your back pain been in the last 2 weeks</td>
<td>No reference standard</td>
<td>No reference standard</td>
</tr>
</tbody>
</table>

*RMDQ* Roland Morris Disability Questionnaire, *TSK* Tampa Scale of Kinesophobia, *HADS* Hospital Anxiety and Depression Scale, *CSQ* Coping Strategy Questionnaire

* Analysis performed on the external patient sample

* Estimates not comparable
standard questionnaires for these three psychosocial constructs. This remains a possibility, as although these translations have been performed, there are no validation studies available for those reference standards. It is also possible that some variability between samples is inevitable, even within the same language population and we do not have data to quantify that variability.

Questions of the acceptability and clinical importance of divergence have been discussed in similar contexts [25]. As the research community has not produced criterion standards to guide decisions on when divergence is clinical important, such results can only be descriptively reported.

The strengths of this study are the staged process of language translation and the direct comparison of discriminative validity with data from the original validation of the English version. A potential weakness of the study is that the translation and validation were not specifically inclusive of non-native Danish speakers as recommended by some methodologists [26, 27]. During the psychometric validation there were no restrictions on the linguistic background of the target population, but it is unknown the extent to which any indirect cross cultural validation may have occurred. The translation and psychometric validation was conducted in the secondary care sector as this was the population of convenience. Subsequent work will investigate its validity in primary care patients and determine whether the cut points in the English version are the most appropriate for the Danish version.

Conclusion

The STarT questionnaire was translated into Danish and its discriminative validity measured. The translation was judged to be linguistically accurate and the STarT-dk tool acceptable for patient use. The discriminative validity largely seemed comparable with the original English version but three psychosocial questions displayed lower discriminative validity. We suspect this is most likely to be a product of differences in severity between the cohorts and variability due to the Danish versions of the reference standard questionnaires not having been validated. Despite those differences, we believe the results show that the STarT-dk has sufficient patient acceptability and discriminative validity to be used in Denmark.

Acknowledgments We are grateful for funding from the Region of Southern Denmark, the Danish Rheumatism Association and the Association of Danish Physiotherapists. We are also grateful to Ms Lene Ververs for data management.

Conflict of interest The authors declare no financial or ethical conflict of interest.

References

8. The STarT Back Screening tool website. 2010. Ref Type: Internet Communication
Manuscript 2:

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

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
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.
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 reported), 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).
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)
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 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

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

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
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**

No data exists on how many were invited to participate in the study. 271 LBP patients were registered in the GP audit.

248 (91%) of the registered patients recruited into the study and completed baseline questionnaires.

204 patients completed and returned their 3-months follow-up questionnaires (72% follow-up rate).

172 patients were randomly selected to match to the number of people in the physiotherapy sample.

The overall cohort from Danish primary care: 172 LBP patients from GP practices + 172 LBP patients from Physiotherapy practices = Total n=344

1. LBP = low back pain.
2. General practitioner.

**References**


Manuscript 3:

Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care?
Original article

Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care?

Lars Morsø*, Peter Kent, Hanne B. Albert, Claus Manniche

Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt, Middelfart, Institute of Regional Health Services Research, University of Southern Denmark, Clinical Locomotion Science Network, Denmark

ARTICLE INFO

Article history:
Received 10 May 2012
Received in revised form 3 July 2012
Accepted 6 July 2012

Keywords: STarT Back Screening Tool Primary care Secondary care Psychosocial profile

ABSTRACT

Differences between the psychosocial risk factors of low back pain (LBP) patients in primary and secondary care are under-investigated. Similarly, differences in the psychosocial profile of people classified into STarT Back Screening Tool (SBT) subgroups in primary and secondary care settings have not been investigated. The aim of the study was to determine: (1) if movement-related fear, catastrophisation, anxiety and/or depression in LBP patients are different between primary and secondary care settings, and (2) if those differences are retained when stratified by SBT subgroup. This study was a cross-sectional comparison of LBP patients in Danish primary settings (405 general practitioner or physiotherapy patients) and a secondary care setting (311 outpatient spine centre patients). Psychosocial factors were measured with the Roland Morris Disability Questionnaire, the Tampa Scale of Kinesiophobia, the Coping Strategies Questionnaire (catastrophisation subscale), and the Hospital Anxiety and Depression Scale. There were significantly higher scores in secondary care for movement-related fear (1.3 points (95%CI 1.1–2.5) p = .030) and catastrophisation (2.0 (95%CI 1.0–3.0) p < .000), lower scores on anxiety (−1.0 (95%CI −1.0–2.0) p < .000) but no difference for depression. These differences in psychosocial scores were broadly retained when stratified by SBT subgroup. However, questionnaire-specific reported thresholds for important difference scores indicate the size of these differences between the care settings were unlikely to be clinically important from a patient perspective. Longitudinal studies are required to investigate the predictive ability of SBT in secondary care settings and whether treatment targeted to SBT subgroups is effective in secondary care.

© 2012 Elsevier Ltd. All rights reserved.

1. Background

Identification of best practice management for low back pain (LBP) is challenged by the heterogeneity of this condition and the difficulty in reaching a definitive diagnosis of a tissue-specific cause for pain in most individual patients (Deyo, 2002; Delitto, 2005; Founrey et al., 2011). International guidelines recommend triaging LBP patients into three broad groups: specific LBP (e.g. ankylosing spondylitis, cancer, and infection), radiculopathy and nonspecific LBP (NSLBP) (Quebec Task Force, 1987; van Tulder et al., 2006; Airaksinen et al., 2006; Krimer and van, 2007; Liddle et al., 2007; Laerum et al., 2007; Dagenais et al., 2010). The group of patients classified with NSLBP represents approximately 75–85% of all primary care patients with LBP (Deyo, 2002). There is considerable diversity in the clinical characteristics of patients within this NSLBP group and there is evidence that this diversity influences outcome (Dunn et al., 2011; Hill and Fritz, 2011).

Guidelines also recommend assessment of the psychosocial characteristics of individual LBP patients, as these have been shown to be risk factors for poor outcome regardless of treatment type (Burton et al., 1995; Linton et al., 2000; Vlaeyen and Linton, 2000; van Tulder et al., 2006; Chou et al., 2007; Krismer and van, 2007; Main et al., 2010; Hill and Fritz, 2011). However, measuring psychosocial factors in the daily clinic can present practical challenges (Kent et al., 2009; Hill et al., 2010b) as many of the questionnaires validated for detecting psychosocial characteristics are comprehensive and time-consuming (Hill et al., 2010a).

There is an increasing focus on the ability of questionnaires to screen LBP patients for psychosocial risk factors in ways that are practical in daily care settings. Examples of such screening questionnaires are the Örebro Musculoskeletal Pain Questionnaire (Linton and Hallden, 1998; Linton and Boersma, 2003) and the STarT Back Screening Tool (SBT) (Hill et al., 2008). The SBT is particularly interesting, for in addition to its prognostic ability, this

Please cite this article in press as: Morsø L et al., Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care?, Manual Therapy (2012), http://dx.doi.org/10.1016/j.math.2012.07.002

* Corresponding author. Research Department, Spine Centre of Southern Denmark, Oestre Houg vej 55, DK-5500 Middelfart, Denmark. Tel.: +45 63484582.
E-mail address: lars.morsoe@slb.regionsyddanmark.dk (L. Morsø)
classification tool also has treatment implications and is very short. It was developed in primary care and is a nine-item multidimensional screening questionnaire that classifies NSLBP patients into subgroups of people with low, medium and high risk of poor prognosis based on the complexity of their clinical profile (Hill et al., 2008). That complexity is based on the physical and psychosocial characteristics of individual patients. In primary care in the United Kingdom, the SBT classification has been shown to be predictive of outcome and there is evidence that treatment matched to these subgroups can be more clinically effective and more cost-effective than usual care (Hill et al., 2011).

However, as the SBT has only been validated in primary care settings, it is not known whether it might also be useful in secondary care settings. This is an important consideration as some authors have suggested that there might be differences in patients’ clinical profiles across care settings (Fritz et al., 2011; Kongsted et al., 2011) and there is a perception that secondary care patients have a more complex psychosocial profile (Pincus et al., 2002). If this were so, then in a generic sense, for prognostic stratification to be useful in secondary care it may need to include different components and similarly, subgroup-stratified treatment could require more complex or intense interventions. In the specific case of the SBT, its classification and predictive ability might be different depending on the care setting.

In Denmark, patients with recent onset of NSLBP are mostly managed in primary care setting by General Practitioners, Physiotherapists or Chiropractors. Danish clinical practice guidelines recommend that NSLBP patients are not referred to specialised secondary care facilities unless an initial course of primary care treatment has been undertaken (Poulsen, 1999). Consequently, patients in Danish secondary care tend to have an extended current pain episode and persistent pain.

Although the influence of psychosocial factors on patient outcomes has been comprehensively investigated in both acute and chronic LBP (Burton et al., 1995; Vlaeyen and Linton, 2000; Pincus et al., 2002; Grotle et al., 2006; Shaw et al., 2009; Hill and Fritz, 2011), differences in the psychosocial profile of LBP patients in different care settings remains under-investigated. For example, it is possible that LBP patients in secondary care might score higher on psychosocial risk factors due to influences such as lengthy episode duration, more persistent pain, non-response to primary care interventions or a different case-mix; however this has not been quantified. Describing such differences would be useful if it provided a better understanding of which clinical characteristics shape the outcomes achievable in each care sector or it identified characteristics suitable for targeting in each setting.

Therefore, in this study we aimed to test the hypothesis that: (1) there would be significant differences between the movement-related fear, catastrophisation, anxiety and/or depression of LBP patients in primary and secondary care settings, and (2) there would be significant differences between care settings on those psychosocial factors when stratified by the SBT into low, medium, high risk classification subgroups.

2. Methods and material

This study is a cross-sectional comparison of the psychosocial profile of patients seeking care for LBP in Danish primary and secondary care settings. Patients with an episode of NSLBP (≥ leg pain) who could read and understand Danish were invited to participate in the study at their first consultation. Participants were asked for their consent and completed a set of questionnaires before their clinical examination. In all settings, participation in the study did not affect usual clinical assessment and care.

The primary care cohort (n = 405) was collected from February to October 2011 at general practitioner (GP) clinics and physiotherapy clinics in the region of Southern Denmark. Questionnaires were given to 271 patients consulting the GP, and consent was given by 205 patients (a 75.6% participation rate) who then completed the questionnaires. Twenty-seven physiotherapy clinics were each sent 10 questionnaire packs to hand out to NSLBP patients seeking care at the physiotherapy clinic, and consent was given by the 200 patients included in this study. No further information about non-responders in primary care is available. The secondary care cohort was collected from May to September 2010 at the Spine Centre of Southern Denmark, which is an outpatient, multi-disciplinary department in a public hospital. In that period, 513 patients were eligible to participate. As 311 consented to the study, the participation rate was 60.6%. The only information available on the 202 non-responders was their age and gender. Non-responders were younger (mean age 48.3) than responders (mean age 53.3), and less likely to be women (non-responders 49.5% women, responders 59.5% women). This study was approved by the Scientific Ethics Committee of the Region of Southern Denmark (S-20100036).

2.1. Data collection

The patient self-reported questionnaire pack contained demographic information, Numerical Pain Rating scales (Manniche et al., 1994), the SBT (Hill et al., 2008), the 23-item version of the Roland Morris Disability Questionnaire (RMDQ) (Roland and Morris, 1983; Albert et al., 2003), the Tampa Scale of Kinesiophobia (TSK) (Swinkels-Meewisse et al., 2003), the Coping Strategies Questionnaire (CSQ) (Rosenstiel and Keefe, 1983), and the Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983). Patients received the questionnaire pack from secretaries in the clinics, who also provided study information and obtained informed consent. All data were double-entered independently into a database (EpiData 3.1, The EpiData Association Odense, Denmark. www.epidata.dk) by two research assistants.

2.2. Sample size

This study is a secondary analysis of data collected for other studies in primary and secondary care. Data from the primary cohort was principally collected for investigating the predictive and external validity of the SBT in Danish primary care (unpublished study). The secondary cohort was principally collected for the translation and discriminative validation of the SBT into Danish (Morso et al., 2011).

A post-hoc sample size calculation was performed based on the available numbers of patients and the standard deviation of the four psychosocial constructs (TSK, CSQ, HADSanx, HADSDep) across care settings, to calculate the smallest amount of difference detectable on each construct. Given these samples and their variance, we had sufficient power to detect a significant difference on the TSK of 1.8; CSQ 1.8; HADSanx 8; HADSDep .8, with a confidence level of 95% and a power of .80.

2.3. Data analysis

Across the collected variables, there was a mean of 3.1% missing data (range 1.7%–9.8%). As it is recommended that when using regression models (Schafer, 1997) missing data should be less than 2%, multiple imputation (van Ginkel and van der Ark, 2012) was conducted. The data distribution was inspected visually and tested for skewness and kurtosis using the methods recommended by Garson (Garson, 2009).
Descriptive statistics (means and SDs, medians and inter-quartile ranges [IQR] and confidence intervals for differences) were used to report the cohort summary characteristics. To test for differences between the primary and secondary care cohorts, we used unpaired t-tests, Mann–Whitney U, Chi square or two-sample Exact Tests of proportions (prtesti) and calculations, two-sample Exact Tests of proportions (prtesti) and confidence intervals of median difference (npshift).

Linear regression was used to determine the contribution of care setting (primary care, secondary care) to the variance in scores on each psychosocial construct (TSK, CSQ, HADSAnx, HADSepep) when controlling for clinically important co-variates (age, gender, episodic duration, back pain, leg pain, activity limitation). Where non-normally distributed, data were log- or square root-transformed to approximate normality, prior to inclusion in regression models. Modelling that entered all variables simultaneously and forward-stepwise regression modelling (p = .5 to enter and p = .10 to remove) were both performed.

The following software packages were used for the statistical procedures. SPSS 13.0 (SPSS Statistics/IBM, Chicago IL, USA) was used for imputation, tests of normality, transformation of data, unpaired t-tests, Mann–Whitney U, Chi square and linear regressions. STATA 11 (STATA corp., 4905 Lakeway Drive, College Station, Texas. 77845 USA) was used for post-hoc sample size calculations, two-sample Exact Tests of proportions (prtesti) and the calculation of confidence intervals of median difference (npshift).

### 3. Results

#### 3.1. Differences between primary and secondary care settings

Overall there was no significant difference in the distribution of the three SBT subgroups in primary and secondary care (p = .246), though there was a slightly higher proportion of medium risk in primary care (42% vs. 32.8%). An unexpectedly high proportion of ‘low risk’ was found among patients referred to secondary care (39.2%).

As seen in Table 1, when comparing the primary and secondary care cohorts, there were more women than men in both cohorts but no significant difference between settings. However, the primary care cohort was significantly younger, more were employed, had a shorter duration of their current LBP episode and had more intense back pain, while leg pain intensity was higher in secondary care. Unexpectedly, both cohorts had near identical levels of activity limitation.

On the psychosocial constructs, there were significantly higher scores in secondary care for movement-related fear and catastrophisation. In contrast, there were lower scores on anxiety in secondary care, but no significant differences for depression. The difference between care settings for movement-related fear was 1.3 (1.1–2.5) and for catastrophisation was 2.0 (1.0–3.0). The scores for anxiety showed a difference of 1.0 (1.0–2.0) but also showed that anxiety was low in absolute terms in both settings, with median values of 4.0 and 5.0 on the 21-point scale.

#### Table 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Primary care (n = 405)</th>
<th>Secondary care (n = 311)</th>
<th>Test for differences between care settings&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female</strong></td>
<td>228 (56.3%)</td>
<td>185 (59.5%)</td>
<td>p = .466</td>
</tr>
<tr>
<td>Missing</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Age mean (±SD years)</td>
<td>49.0 (14.0)</td>
<td>53.3 (15.7)</td>
<td>p &lt; .000</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Employed, proportion&lt;sup&gt;b&lt;/sup&gt;</td>
<td>265 (65.4%)</td>
<td>149 (47.9%)</td>
<td>p &lt; .000</td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Duration of episode, weeks&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 0–2</td>
<td>119 (29.4%)</td>
<td>12 (3.9%)</td>
<td>p &lt; .000</td>
</tr>
<tr>
<td>• 2–4</td>
<td>58 (14.3%)</td>
<td>9 (2.9%)</td>
<td>p &lt; .000</td>
</tr>
<tr>
<td>• 4–12</td>
<td>75 (18.5%)</td>
<td>43 (13.8%)</td>
<td>p = .055</td>
</tr>
<tr>
<td>• &gt;12</td>
<td>143 (35.3%)</td>
<td>187 (60.1%)</td>
<td>p &lt; .000</td>
</tr>
<tr>
<td>Missing</td>
<td>10</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>12</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Leg pain (scale 0–10)</td>
<td>3.0 [0–6]</td>
<td>4.0 [1–7]</td>
<td>p = .017</td>
</tr>
<tr>
<td>Missing</td>
<td>14</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>SBT Back Tool group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low risk</td>
<td>140 (34.6%)</td>
<td>122 (39.2%)</td>
<td>p = .246</td>
</tr>
<tr>
<td>• Medium risk</td>
<td>170 (42.0%)</td>
<td>102 (32.8%)</td>
<td></td>
</tr>
<tr>
<td>• High risk</td>
<td>95 (23.3%)</td>
<td>86 (27.7%)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Activity limitation&lt;sup&gt;d,e&lt;/sup&gt;  (scale 0–100)</td>
<td>60.8 [39–74]</td>
<td>60.9 [39–78]</td>
<td>p = .678</td>
</tr>
<tr>
<td>Median [iqr]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear of movement&lt;sup&gt;d&lt;/sup&gt;       (scale 17–68)</td>
<td>35.8 (7.6)</td>
<td>37.1 (8.5)</td>
<td>p = .030</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catastrophisation&lt;sup&gt;d&lt;/sup&gt;      (scale 0–36)</td>
<td>10.0 [5–15]</td>
<td>12.0 [7–18]</td>
<td>p &lt; .000</td>
</tr>
<tr>
<td>Median [iqr]</td>
<td>5.0 [3–8]</td>
<td>4.0 [1–6]</td>
<td>p &lt; .000</td>
</tr>
<tr>
<td>Anxiety&lt;sup&gt;d&lt;/sup&gt;                (scale 0–21)</td>
<td>2.0 [1–5]</td>
<td>2.0 [0–5]</td>
<td>p = .446</td>
</tr>
<tr>
<td>Depression&lt;sup&gt;d&lt;/sup&gt;             (scale 0–21)</td>
<td>2.0 [1–5]</td>
<td>2.0 [0–5]</td>
<td>p = .446</td>
</tr>
</tbody>
</table>

<i>qqr = interquartile range.</i>

<sup>a</sup> Primary care cohort from general practice and physiotherapist clinics, secondary care cohort from a hospital outpatient department.

<sup>b</sup> Employed or being student.

<sup>c</sup> Current episode.

<sup>d</sup> RMDQ = Roland Morris Disability Questionnaire, TSK = Tampa Scale of Kinesophobia, CSQ = Coping Strategy Questionnaire (catastrophisation subscale) HADS = Hospital Anxiety and Depression Scale.

<sup>e</sup> Imputed data used for analysis.

<sup>f</sup> T-test, Mann–Whitney U or Chi square used, depending on the data and distribution.
The results of the standard linear regression models for each psychosocial construct are shown in Table 2. In each of those models, one of the psychosocial constructs was the dependent variable and the following variables were all entered simultaneously as independent predictor variables: care setting, age, gender, employment status, pain intensity, activity limitation and duration of complaint. Overall, 13%–31% of the total variance of each psychosocial construct was explained by the entered variables. In these models, the beta-coefficients for care-setting were only statistically significant for catastrophisation ($b = .104, p = .006$) and anxiety ($b = .101, p = .020$). When regression models were repeated using a forward stepwise procedure, for the models of depression and movement-related fear, the variable of care setting was not selected for inclusion at any stage in the model building. For the stepwise modelling model for anxiety, care setting was entered in the 4th of 5 stages, and for the modelling of catastrophisation it was entered in the 6th of 7 stages (results not shown). These stepwise results collectively indicate that care setting was not a dominant influence on the way people scored these psychosocial constructs.

3.2. Differences between primary and secondary care settings when stratified by StartT back screening tool subgroups

The statistically significant differences in psychosocial scores that were observed between care settings were broadly retained when stratified by SBT subgroup (Fig. 1). In SBT subgroups 2 and 3, there was more movement-related fear and catastrophisation in secondary care. In contrast, in all SBT subgroups there was less anxiety in secondary care.

In both care settings, there was a trend towards increasing psychosocial scores on all four psychosocial constructs across SBT subgroups from low-risk to high-risk, indicating that the stratification in the primary care setting where SBT has previously been validated, was mirrored in the secondary care setting.

4. Discussion

This study tested the hypotheses that: (1) there would be significant differences between the movement-related fear, catastrophisation, anxiety and/or depression of LBP patients in primary and secondary care settings, and (2) there would be significant differences between care settings on those psychosocial factors when stratified by SBT classification subgroup (low, medium, high risk subgroups). The results support both hypotheses, showing small but statistically significant differences between care settings on three of the four psychosocial constructs that we tested, and those differences were retained when stratified by SBT subgroup. However, although these differences were statistically significant, we do not consider them to be clinically important in size. That is because a score of 4 TSK points or more has been reported by Woby et al. (Woby et al., 2005) as most accurately indicating a clinically important difference in fear of movement and so the 1.3 TSK points difference between our care settings was unlikely to be clinically important in size. Similarly, 1.7 CSQ catastrophising subscale points and 1.4 HADS anxiety points were reported by Angst et al. (Angst et al., 2008) as minimal clinically important differences (MCID), and our differences were 2.0 (catastrophisation) and 1.0 (anxiety). But they note that their distribution-based method for estimating MCID is ‘much more conservative’ than anchor-based methods that ask patients to rate their perceived change (Angst et al., 2008).

Therefore in our view, the threshold values reported in those studies suggest that the differences between the care settings seen in our study were unlikely to be clinically important from a patient perspective.

In the regression models, the variables of care setting, age, gender, pain intensity, activity limitation and duration of pain episode collectively explained a limited amount (13%–31%) of the variance in the included psychosocial measures. In half the standard regression models, care setting was not a statistically significant contributor, and the infrequent and late-stage inclusion of care setting in step-wise regression models also indicate that the contribution of care setting to the psychosocial scores was minor.

The distribution across the three SBT subgroups in primary care was very similar to that reported by Hill et al. (Hill et al., 2008) in a primary care study in the UK (low risk 40%, medium risk 35% and high risk 25%). This reinforces the findings from the Danish SBT translation study, where the construct validity of the Danish version of the SBT was found to be in concordance with the original SBT (Morso et al., 2011). Stratification by SBT subgroup showed increasingly higher scores across SBT subgroups on each psychosocial construct. As this pattern was consistent across care settings, this also reinforces the construct validity of the SBT. This may

<table>
<thead>
<tr>
<th>Fear of movement$^d$</th>
<th>Catastrophisation$^d$</th>
<th>Anxiety$^d$</th>
<th>Depression$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-coefficient$^e$ (CI 95%)</td>
<td>p-value$^f$</td>
<td>Beta-coefficient$^e$ (CI 95%)</td>
<td>p-value$^f$</td>
</tr>
<tr>
<td>Care setting$^g$</td>
<td>Age</td>
<td>Gender</td>
<td>Employment status$^h$</td>
</tr>
<tr>
<td>.036 (−2.6 to 2.6)</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>.357</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>.104 (−1.0 to 1.2)</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>.006</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>.101 (−1.1 to 1.3)</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>.020</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>.062 (−2 to 3)</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>.141</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
</tbody>
</table>

$^a$ Proportion employed, defined as either being employed or a student.  
$^b$ Duration of low back pain episode in weeks.  
$^c$ Numeric range scale (0–10).  
$^d$ Roland Morris Disability Questionnaire (0–100 proportional scale), Tampa Scale of Kinesophobia (17–68 scale), Coping Strategy Questionnaire (0–36 scale of the catastrophisation subscale), Hospital Anxiety and Depression Scale (0–21 scale).  
$^e$ Primary care cohort from general practice and physiotherapist clinics. Secondary care cohort from a specialised outpatient spine centre.  
$^f$ Standardised Beta coefficient.  
$^g$ Significance level p < .05.
indicate that the SBT can accurately classify people in the secondary care setting, though the prognostic and therapeutic implications of such classification are untested.

Although patients referred to Danish secondary care usually have received an initial course of primary care treatment, a large proportion of those referred to our secondary care setting were SBT ‘low risk’ patients. This seems to be a contradiction, as people classified into the SBT low-risk category are expected to recover well in primary care. There could be many reasons for this. One reason could be that some low-risk patients fail to improve in primary care due to iatrogenic factors, such as inadequate reassurance and information on self-management or over-treating. Another reason could be that the SBT does not detect important clinical characteristics in patients that are associated with persistent LBP in secondary care. For example, family or work-related factors or other characteristics from the full biopsychosocial spectrum might also be important in this phase of an episode.

In this study, there were small average differences between care settings on specific psychosocial factors, albeit that they were unlikely to be clinically important in size. We could speculate that these differences might reflect aspects of the transition from sub-acute to persistent pain. For example, it could be that anxiety is a more prevalent adaptive response in recently-onset pain but this transition to catastrophisation or fear of movement as the pain persists over time despite reassurance and treatment. Investigation of this would require detailed longitudinal tracking of inception cohorts.

The use of the SBT as a way to target treatment in primary care has shown promising results (Hill et al., 2011), in addition to its use as a prognostic classification tool. Targeting treatment to specific subgroups of primary care patients based on their SBT classification resulted in better patient outcomes and was more cost effective than usual physiotherapy care in a recent randomised controlled trial (Hill et al., 2011). Maybe psychologically-informed physiotherapy treatment targeted to ‘high risk’ patients in primary care would reduce the number of referrals required to secondary care.

Possibly this treatment approach has a role for ‘high risk’ patients in secondary care or alternatively, it could be that more intensive psychological support is required in this setting. However, investigation of the predictive ability and effectiveness of treatment targeted to SBT subgroups in secondary care was beyond the scope of our study as only cross-sectional data were available.

A strength of the current study is that data from 711 patients were available from most types of care settings that are common for back pain patients in our health region. However, the notable exception is chiropractic patients. The data collection also covered GPs and physiotherapists across a broad geographic area within our health region and included its major spine centre. Therefore, the results of the study are likely to have good generalisability within that region and maybe more broadly within Denmark. Their generalisability internationally would require testing in further studies. Potential weaknesses of the data are that although recruitment was designed to be sequential consenting patients, no data were available as to the adherence to this practice in the primary care recruitment, and that minimal information is available about non-participant bias.

5. Conclusion

LBP patients in secondary care had movement-related fear and catastrophisation scores that were higher and anxiety scores that were lower than those in primary care, at a statistically significant level. When stratified by SBT risk subgroup, patients in secondary care in the medium risk and high risk subgroups had significantly higher movement-related fear, catastrophisation and depression, but in the low and medium risk groups had lower anxiety scores than those in primary care. Though statistically significant, the size of these differences was unlikely to be clinically important from a patient perspective.

Stratification by SBT subgroup showed increasingly higher scores on each psychosocial construct across groups and this reinforces the construct validity of the SBT in both care settings.
Longitudinal studies are required to investigate the predictive ability of the SBT in secondary care settings and whether treatment targeted to SBT sub-groups is also effective in secondary care.

Acknowledgements

The authors thank Lene Ververs, Jytte Johansen and Orla Lund Nielsen for assistance in handling the questionnaires. We would also like to thank Lise Hestbaek and Alice Kongsted for support with collecting data from the GPs and to physiotherapists from primary care in the Region of Southern Denmark for collecting data in their clinics.

References


Available from: PM:29610863.


Available from: PM:29610863.


Manuscript 4:

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

Lars Morsø¹, Peter Kent¹, Hanne B Albert¹, Claus Manniche¹

¹Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt, Middelfart, Institute of Regional Health Services Research, University of Southern Denmark, a member of the Clinical Locomotion Network.

Corresponding author:
Lars Morsø, Research Department, Spine Centre of Southern Denmark, Oestre Hougvej 55 DK-5500 Middelfart, phone: +4563484582, e-mail: lars.morsoe@slb.regionsyddanmark.dk
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 compare the predictive ability of the SBT in Danish secondary and primary care settings.

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 (AUC) 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. Investigation of 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 nonspecific 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 have 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 randomized 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 general, it is recommended that external validation of questionnaires across care settings should be undertaken due to differences in patient 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: (i) its predictive ability (prognostic accuracy), and (ii) 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 are referred to the Spine Centre by GPs, chiropractors or medical specialists for a comprehensive evaluation due to suboptimal improvement during assessment in primary care. 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 some 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. 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 twenty-seven 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 data from secondary care was
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)
was 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 to 10 Numerical Rating Scales). 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 6-month follow-up. The cut point used in 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, that 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 2x2 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 people without the following outcomes at 6 months follow-up: (i) activity limitation as defined above, and (ii) 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 different between the cohorts. All covariates were initially entered into the model, followed by a manual backwards stepwise reduction (p<.05 to remove) to the most parsimonious model. An odds ratio greater that 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 in order to have a predictive reference standard to compare those results to, post-hoc we also calculated 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 to 10 pain intensity scale. The same distribution-based method was used for categorising the baseline activity limitation scores on the 0 to 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<.001), leg pain intensity (p<.001) and borderline significant on LBP intensity (p=.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 (IQR65-87) in secondary care and 77.8 (IQR70-84) in primary care (0 to 100 scale).
**TABLE 1 ABOUT HERE**

*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 \((p<.001)\) (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 2 ABOUT HERE**

*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 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.

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 (Figure 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 is not as strong in secondary care as it was in primary care (RR medium risk 2.3 [CI95% 1.2; 4.5], high risk 3.5 [CI95% 1.8; 6.6].

It is likely that these two findings of: (i) nearly half the low risk subgroup in secondary care having a poor outcome and (ii) 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.

FIGURE 1 ABOUT HERE

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

TABLE 3 ABOUT HERE

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 (-.005 in secondary care and .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, baseline episode duration was predictive in both cohorts and had an influence that was independent of the predictive ability of the SBT subgroups. The same was true for baseline LBP intensity in secondary care.

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 RR for baseline pain intensity were medium risk 1.5[1.3; 1.7], high risk 1.6[1.4; 1.8] and for activity limitation were medium risk 1.6[1.4; 1.8], 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 people 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.

TABLE 4 ABOUT HERE
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], the 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 risks 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 categorized baseline pain and activity limitation were no stronger at prediction in this cohort, despite it being well recognised that these are strong predictors.
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 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 sociodemographics in the secondary care cohort. In pregnancy-related pelvic pain it has been shown that sociodemographics 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% to 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 are referred to as ‘low complexity’ and 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 sumscores were used as a continuous scale in longitudinal modelling of a non-uniform outcome period [8]. In the other study the SBT sumscores 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, this was investigated by us in secondary care where care pathways were uninfluenced by 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. 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, episode duration was a predictive factor that was independent of SBT subgroup classification, and baseline pain intensity also was in secondary care. The usefulness of SBT subgroup targeted treatment in secondary care was not investigated in this study.
Reference List


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

<table>
<thead>
<tr>
<th>Secondary care cohort</th>
<th>Total cohort (n=960)</th>
<th>SBT Low risk n=252 (27.7%)</th>
<th>SBT Medium risk n=296 (32.5%)</th>
<th>SBT High risk n=363 (39.9%)</th>
<th>Test for differences between SBT subgroups*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age in years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mean, (SD)</td>
<td>52.0 (14.1)</td>
<td>52.3 (14.6)</td>
<td>52.0 (13.8)</td>
<td>51.3 (14.0)</td>
<td>p=.673</td>
</tr>
<tr>
<td><strong>Female, proportion</strong></td>
<td>521 (54.3%)</td>
<td>116 (46%)</td>
<td>179 (60.5%)</td>
<td>194 (53.4%)</td>
<td>p=.003</td>
</tr>
<tr>
<td><strong>Duration in months</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• &lt;1 month</td>
<td>47 (5.0%)</td>
<td>5 (2.0%)</td>
<td>18 (6.3%)</td>
<td>22 (6.3%)</td>
<td>p=.012</td>
</tr>
<tr>
<td>• 1 to 3 months</td>
<td>139 (14.9%)</td>
<td>36 (14.6%)</td>
<td>56 (19.6%)</td>
<td>41 (11.6%)</td>
<td></td>
</tr>
<tr>
<td>• &gt; 3 months</td>
<td>746 (80.0%)</td>
<td>205 (83.3%)</td>
<td>212 (74.1%)</td>
<td>289 (82.1%)</td>
<td></td>
</tr>
<tr>
<td><strong>Pain intensity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low back pain</td>
<td>6 (4-7)</td>
<td>4.3 (2-6)</td>
<td>6.0 (5-7)</td>
<td>7.0 (6-8)</td>
<td>p&lt; .001</td>
</tr>
<tr>
<td>• Leg pain</td>
<td>5 (2-7)</td>
<td>3.0 (1-5)</td>
<td>5.0 (3-6)</td>
<td>6.3 (4-8)</td>
<td>p&lt; .001</td>
</tr>
<tr>
<td><strong>Activity limitation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Median</td>
<td>60.9 (44-78)</td>
<td>34.8 (22-52)</td>
<td>60.9 (48-74)</td>
<td>78.3 (65-87)</td>
<td>p&lt; .001</td>
</tr>
<tr>
<td>• Proportion of patients &gt;30</td>
<td>838 (88.6%)</td>
<td>158 (64.2%)</td>
<td>281 (95.9%)</td>
<td>354 (98.9%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary care cohort</th>
<th>Total cohort (n=172)</th>
<th>SBT Low risk n=54 (34.6%)</th>
<th>SBT Medium risk n=64 (41.0%)</th>
<th>SBT High risk n=38 (24.4%)</th>
<th>Test for differences between SBT subgroups*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age in years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mean, (SD)</td>
<td>52.0 (15.2)</td>
<td>53.3 (14.6)</td>
<td>50.4 (14.6)</td>
<td>49.5 (16.7)</td>
<td>p=.377</td>
</tr>
<tr>
<td><strong>Female, proportion</strong></td>
<td>98 (57.0%)</td>
<td>30 (55.6%)</td>
<td>41 (64.1%)</td>
<td>17 (44.7%)</td>
<td>p=.163</td>
</tr>
<tr>
<td><strong>Duration in months, proportion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• &lt;1 month</td>
<td>65 (38.9%)</td>
<td>18 (34.0%)</td>
<td>29 (47.5%)</td>
<td>13 (34.2%)</td>
<td>p=.246</td>
</tr>
<tr>
<td>• 1 to 3 months</td>
<td>39 (23.4%)</td>
<td>17 (32.1%)</td>
<td>13 (21.3%)</td>
<td>7 (18.4%)</td>
<td></td>
</tr>
<tr>
<td>• &gt; 3 months</td>
<td>63 (37.7%)</td>
<td>18 (34.0%)</td>
<td>19 (31.1%)</td>
<td>18 (47.7%)</td>
<td></td>
</tr>
<tr>
<td><strong>Pain intensity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low back pain</td>
<td>5.0 (4-7)</td>
<td>4.0 (3-6)</td>
<td>6.0 (4-8)</td>
<td>6.0 (5-7)</td>
<td>p&lt; .001</td>
</tr>
<tr>
<td>• Leg pain</td>
<td>3.0 (0-6)</td>
<td>1.5 (0-3)</td>
<td>3.0 (1-6)</td>
<td>4 (1-7)</td>
<td>p&lt; .001</td>
</tr>
<tr>
<td><strong>Activity limitation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Median</td>
<td>60.9 (39-77)</td>
<td>34.8 (26-49)</td>
<td>67.4 (52-78)</td>
<td>77.8 (70-84)</td>
<td>p&lt; .001</td>
</tr>
<tr>
<td>• Proportion of patients &gt;30</td>
<td>146 (86.3%)</td>
<td>36 (66.7%)</td>
<td>60 (93.8%)</td>
<td>38 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

1Standard deviation
2Inter-quartile range
3Numeric Rating Scale (0-10)
High scores are worse
Roland Morris Disability Questionnaire
Tests for differences between subgroups were ANOVA or Kruskal Wallis procedures, depending on data type and distribution
Table 2. Outcome at 6 months follow-up for the Danish secondary and primary care cohorts.

<table>
<thead>
<tr>
<th>Secondary care cohort</th>
<th>Total cohort n=960</th>
<th>SBT Low risk n=252 (27.7%)</th>
<th>SBT Medium risk n=296 (32.5%)</th>
<th>SBT High risk n=363 (39.9%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain intensity¹ (0-10 scale)²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low back pain, median, (IQR³)</td>
<td>4.7 (2-6)</td>
<td>3.0 (2-5)</td>
<td>4.3 (2-6)</td>
<td>5.7 (3-7)</td>
</tr>
<tr>
<td>• Leg pain, median, (IQR³)</td>
<td>2.7 (0-5)</td>
<td>1.3 (0-4)</td>
<td>2.7 (0-5)</td>
<td>3.7 (1-6)</td>
</tr>
<tr>
<td>Activity limitation⁴ (0-100 scale)²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Median, (IQR³)</td>
<td>47.8 (28-70)</td>
<td>26.1 (13-48)</td>
<td>47.8 (22-70)</td>
<td>65.2 (35-83)</td>
</tr>
<tr>
<td>• Proportion of patients &gt; 30⁵</td>
<td>656 (69.0%)</td>
<td>120 (47.8%)</td>
<td>209 (71.3%)</td>
<td>292 (81.3%)</td>
</tr>
<tr>
<td>Primary care cohort</td>
<td>Total cohort n=144</td>
<td>SBT Low risk n=48 (36.9%)</td>
<td>SBT Medium risk n=52 (40.0%)</td>
<td>SBT High risk n=30 (23.1%)</td>
</tr>
<tr>
<td>Pain intensity¹ (0-10 scale)²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low back pain, median, (IQR³)</td>
<td>3.0 (2-5)</td>
<td>3.0 (2-4)</td>
<td>2.0 (0-4)</td>
<td>4.5 (2-5)</td>
</tr>
<tr>
<td>• Leg pain, median, (IQR³)</td>
<td>1.3 (0-4)</td>
<td>0.5 (0-3)</td>
<td>2.0 (1-5)</td>
<td>1.0 (0-5)</td>
</tr>
<tr>
<td>Activity limitation⁴ (0-100 scale)²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Median, (IQR³)</td>
<td>26.0 (9-57)</td>
<td>17.0 (4-26)</td>
<td>30.0 (9-55)</td>
<td>65.0 (25-75)</td>
</tr>
<tr>
<td>• Proportion of patients &gt; 30⁵</td>
<td>51 (40.2%)</td>
<td>9 (20.0%)</td>
<td>21 (46.7%)</td>
<td>18 (69.2%)</td>
</tr>
</tbody>
</table>

¹ Numeric Rating Scale
² High scores are worse
³ Inter-quartile range
⁴ Roland Morris Disability Questionnaire
⁵ Patients with a Roland Morris Disability Questionnaire score >30 were classified as having persistent activity limitation
Table 3. The odds of having a poor clinical outcome on activity limitation\(^1\) at 6 months follow-up in the Danish secondary care and primary care cohorts, estimated by STarT Back Tool subgroup using logistic regression.

<table>
<thead>
<tr>
<th></th>
<th>Secondary care cohort (n=903)</th>
<th>Primary care cohort (n=116)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio [CI 95%]</td>
<td>Odds ratio [CI 95%]</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>Unadjusted model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STarT Back Tool low-risk subgroup(^2)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>STarT Back Tool medium-risk subgroup</td>
<td><strong>2.72</strong> [1.91; 3.87]</td>
<td><strong>3.50</strong> [1.37; 8.93]</td>
</tr>
<tr>
<td>STarT Back Tool high-risk</td>
<td><strong>4.76</strong> [3.31; 6.84]</td>
<td><strong>9.00</strong> [2.97; 27.25]</td>
</tr>
<tr>
<td>Constant</td>
<td>.92 [0.72; 1.17]</td>
<td>.25 [0.12; 0.52]</td>
</tr>
<tr>
<td><strong>Model adjusted episode duration and baseline low back pain</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STarT Back Tool low-risk subgroup(^2)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>STarT Back Tool medium-risk subgroup</td>
<td><strong>2.36</strong> [1.61; 3.47]</td>
<td><strong>4.68</strong> [1.66; 13.18]</td>
</tr>
<tr>
<td>STarT Back Tool high-risk</td>
<td><strong>3.31</strong> [2.18; 5.03]</td>
<td><strong>9.13</strong> [2.60; 32.01]</td>
</tr>
<tr>
<td><strong>Episode duration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1 month</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1-3 months</td>
<td>1.08 [0.52; 2.23]</td>
<td>4.38 [1.19; 16.05]</td>
</tr>
<tr>
<td>&gt; 3 months</td>
<td>2.10 [1.10; 4.02]</td>
<td>6.45 [2.20; 18.84]</td>
</tr>
<tr>
<td><strong>Low back pain intensity baseline</strong></td>
<td>1.17 [1.09; 1.26]</td>
<td>1.00</td>
</tr>
<tr>
<td>Constant</td>
<td>.252 [0.12; 0.53]</td>
<td>.001</td>
</tr>
</tbody>
</table>

\(^1\)Roland Morris Disability Questionnaire score > 30 (0 to 100 scale)

\(^2\)Reference category
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.

<table>
<thead>
<tr>
<th>Outcome Description</th>
<th>Danish secondary care cohort AUC(^1) [95%CI]</th>
<th>Danish primary care cohort AUC(^1) [95%CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>People with a Roland Morris Disability Questionnaire score &gt; 30 (0-100 scale) at 6 months</td>
<td>.69 [.66; .73]</td>
<td>.73 [.64; .82]</td>
</tr>
<tr>
<td>People with severe back pain at 6 months (8-10 on a 0-10 scale)</td>
<td>.72 [.68; .77]</td>
<td>.66 [.46; .85]</td>
</tr>
</tbody>
</table>

\(^1\)AUC = the Area Under the Curve statistic from Receiver Operating Characteristic (ROC) curves
Figure 1. Relative risk of poor clinical outcome\(^1\) on activity limitation at 6 months by SBT subgroup in the Danish secondary and primary care cohorts.

\(^1\)More than 30 Roland Morris Disability Questionnaire points on a recalculated 0-100 scale
\(*P<.05\)
PhD thesis - Appendices

Appendix 1: Scoring of the SBT

Appendix 2: The validated version of the Danish translated SBT
Appendix 1.

Score flow of the SBT
Appendix 2.

The validated version of the translated SBT

STærT Spørgeskemaet

Patientens navn: _______________________________    Dato: _____________

Tænk tilbage på de seneste 2 uger og marker dit svar på følgende spørgsmål:

<table>
<thead>
<tr>
<th>Spørgsmål</th>
<th>Nej</th>
<th>Ja</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I løbet af de seneste 2 uger har mine rygsmerter bredt sig ned i mit/mine ben</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Jeg har haft smerter i mine skuldre eller nakke i løbet af de seneste 2 uger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Jeg har kun gået korte afstande på grund af mine rygsmerter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I løbet af de seneste 2 uger har jeg klædt mig langsommere på end normalt på grund af rygsmerter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Det er egentligt ikke sikkert for en person i min tilstand at være fysisk aktiv</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Jeg har været bekymret meget af tiden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Jeg føler mine rygsmerter er forfærdelige og de bliver aldrig bedre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Generelt har jeg ikke nydt alle de ting, som jeg plejede at nyde</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Overordnet set, hvor generende har dine rygsmerter været de seneste 2 uger?

<table>
<thead>
<tr>
<th>Slet ikke</th>
<th>Ligt</th>
<th>Middel</th>
<th>Meget</th>
<th>Extremt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total score (alle 9): ________________    Sub Score (spr. 5-9):______________