**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*

## 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.
   Morsø L, Albert H, Kent P, Manniche C, Hill J. *Eur Spine J*. 2011 Dec;20(12):2166-73.

2. The predictive and external validity of the STarT Back Tool in Danish primary care.
   Morsø L, Kent P, Albert HB, Hill JC, Kongsted A, Manniche C. *Eur Spine J*. 2013 Feb 10. [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?
   Morsø L, Kent P, Albert HB, Manniche C. *Man Ther*. 2013 Feb;18(1):54-9.

4. The predictive ability of the STarT Back Screening tool in a Danish secondary care setting.
   Morsø L, Kent P, Albert HB, Manniche C. (Submitted to *Eur Spine J*. 18. February 2013)

## 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

STarT Back Screening Tool (SBT) er et 9-punkts patientspørgeskema til inddeling af patienter med uspecifikke lænderyg smerter. Dette korte multidimensionelle spørgeskema er udviklet til primær sektor og kan identificere modificérbare risiko faktorer som udbredt smerte, funktionsnedsættelse og psykosociale elementer. Patienterne inddeles i 3 subgrupper (lav, mellem eller høj) alt efter risiko for dårlig prognose. Det er vist, at SBT har både prognostisk og behandlingsmæssig implikation. Det aktuelle Ph.d. projekt havde til mål at: (i) Oversætte SBT til dansk, (ii) Teste den interne validitet, (iii) Teste den prædiktive validitet i dansk primær sektor, (iv) Undersøge forskelle i den psykosociale patient profil mellem dansk primær- og sekundærsektor, (v) Undersøge den prædiktive evne i dansk sekundær sektor.

### 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

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?** |
|---|
| • **SBT kan identificere modificerbare risiko faktorer i primær sektor.** |
| • **SBT kan klassificere patienter i relevante subgrupper** |
| • **SBT har prognostisk og behandlingsmæssig implikation.** |
| • **Målrettet SBT behandling har vist klinisk og økonomisk effekt.** |

| **Hvad har dette Ph.d. projekt bidraget med?** |
|---|
| • **Oversættelsen af SBT til dansk er forståelig og anvendelig.** |
| • **Den diskriminative validitet af SBT acceptabel** |
| • **Den prædiktive værdi i primær sektor er acceptabel.** |
| • **Prædiktion af prognose i sekundær sektor reduceret.** |

## Konklusion

Samlet set viste resultaterne fra oversættelsen, valideringen og test af den prædiktiv evne, at SBT er et anvendeligt og brugbart klassifiseringsredskab i dansk primær sektor. På trods af sammenlignelige psykosociale patient profiler på tværs 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.

| **What was known prior to this PhD project?** |
|---|
| • **The SBT can identify modifiable risk factors in primary care.** |
| • **The SBT can classify patients into relevant subgroups.** |
| • **The SBT has prognostic and treatment implications.** |
| • **The targeting of treatment has been shown to have positive clinical effectiveness and economic impact.** |

## 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

| **What does this PhD project add?** |
|---|
| • **The Danish translation of SBT was linguistically accurate and acceptable to patients.** |
| • **The discriminative validity of SBT was acceptable.** |
| • **The predictive ability in primary care was acceptable.** |
| • **The predictive ability of SBT in secondary care was less.** |

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:

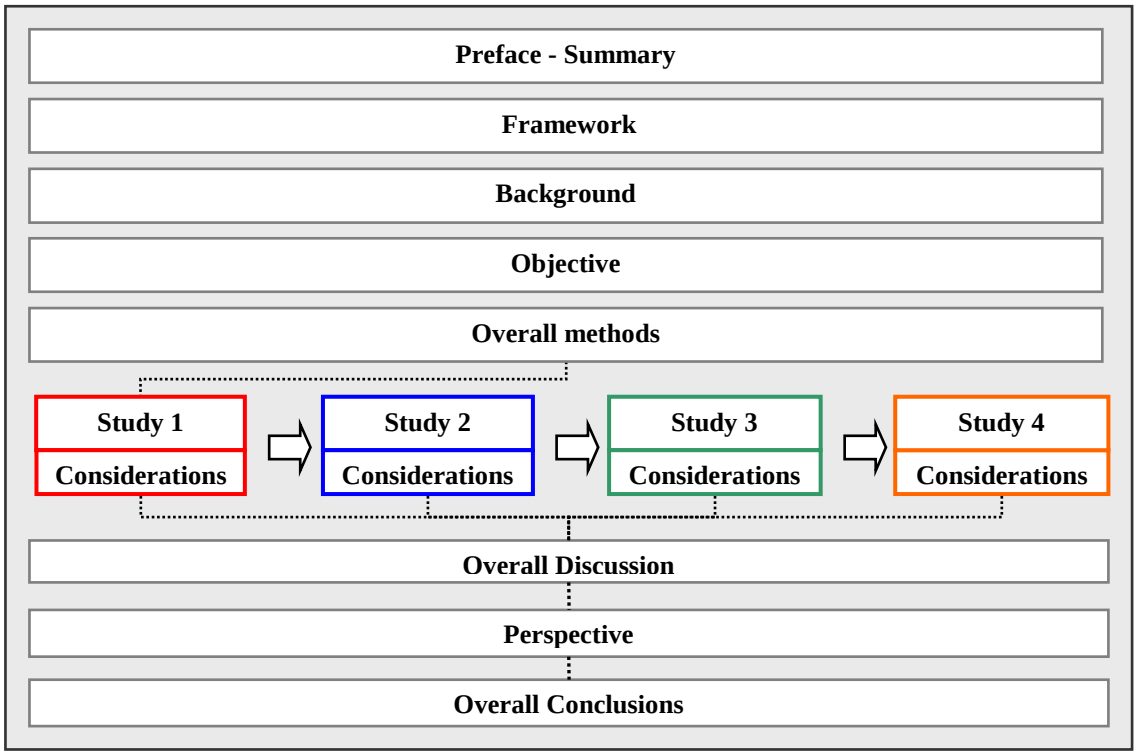



Figure 1. Flowchart of the thesis

## *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

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 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

| | Danish translation cohort from secondary care | Danish primary cohort (GP, PT) | Danish secondary predictive cohort | UK primary care development cohort | UK primary care BeBack cohort |
|---|---|---|---|---|---|
| **Study 1** | Baseline only | | | Baseline only | |
| **Study 2** | | Baseline & 3-month outcomes | | | Baseline & 3-month outcomes |
| **Study 3** | Baseline only | Baseline only | | | |
| **Study 4** | | Baseline & 6-month outcomes | Baseline & 6-month outcomes | | |

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

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

| Danish translation cohort from secondary care | Danish primary cohort (GP, PT) | Danish secondary predictive cohort | UK primary care development cohort | UK primary care BeBack cohort |
|---|---|---|---|---|
| Baseline only | | | Baseline only | |

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

| Phases | **Tasks of the translation process** |
|---|---|
| 1 | Liaison with SBT developers |
| 2 | Translation from English to Danish |
| 3 | Back translation from Danish to English |
| 4 | Synthesis |
| 5 | Translation committee consensus? |
| 6 | Pilot testing |
| 7 | Testing of final version |

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)

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

## 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

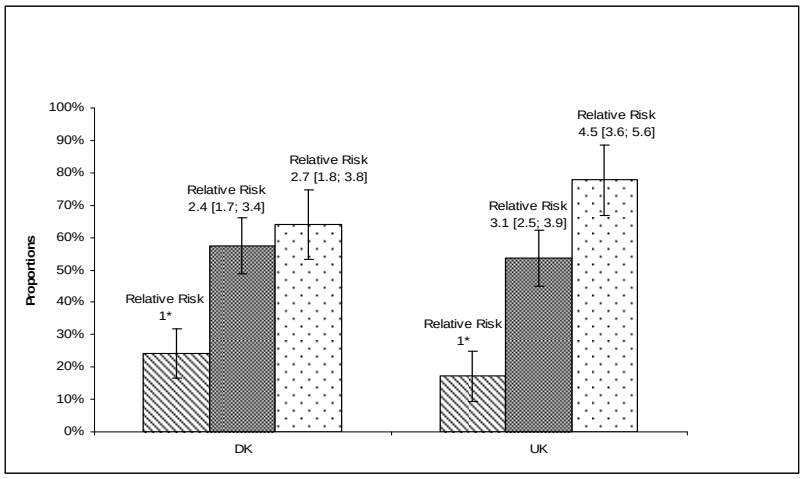
| Danish translation cohort from secondary care | Danish primary cohort (GP, PT) | Danish secondary predictive cohort | UK primary care development cohort | UK primary care BeBack cohort |
|---|---|---|---|---|
| | Baseline & 3 month-outcomes | | | Baseline & 3 month-outcomes |

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).

Figure 2. Proportions of patients within each SBT subgroup who had a poor clinical outcome on *activity limitation* at 3 months and their relative risk.




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

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)

| | Danish cohort (n=322) | | UK cohort (n=845) | |
|---|---|---|---|---|
| | Odds ratio [CI 95%] | p-value | Odds ratio [CI 95%] | p-value |
| ***Unadjusted model*** | | | | |
| SBT low-risk subgroup[2] | 1.00 | | 1.00 | |
| SBT medium-risk subgroup | **4.24** [2.45; 7.32] | p<.001 | **5.56** [3.99; 7.76] | p<.001 |
| SBT high-risk subgroup | **5.57** [2.97; 10.47] | p<.001 | **16.88** [9.71; 29.34] | p<.001 |
| Constant | .32 [.21; .48] | p<.001 | .22 [.16; .26] | p<.001 |
| ***Parsimonious model adjusted for care setting and change on SBT psychosocial constructs (n=296), using manual backwards step-wise procedure*** | | | | |
| SBT low-risk subgroup[2] | 1.00 | | | |
| SBT medium-risk subgroup | **7.89** [3.87; 16.11] | p<.001 | | |
| SBT high-risk subgroup | **15.73** [6.60; 37.47] | p<.001 | | |
| Care setting[3] | 0.31 [0.13; .71] | p=.006 | | |
| Change in anxiety[6] | 0.81 [0.73; .89] | p<.001 | | |
| Change in pain bothersomeness[7] | 0.27 [0.17; .43] | p<.001 | | |
| Interaction between care setting and change in pain bothersomeness | 2.48 [1.41; 4.34] | p=.002 | | |
| Constant | 1.02 [0.49; 2.15] | p=.951 | | |

## 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,

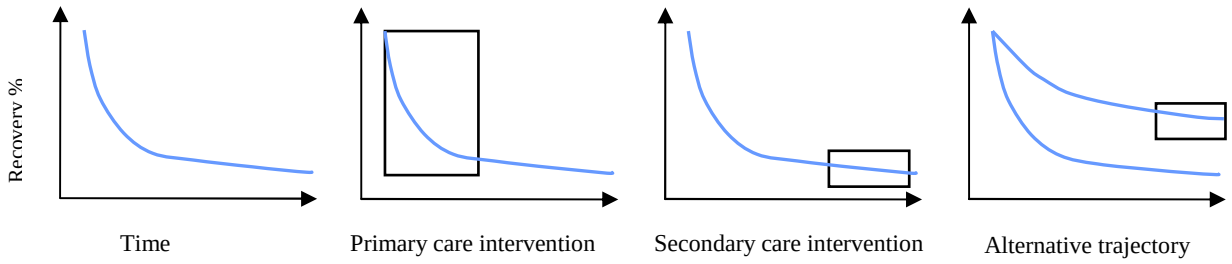
the predictive ability of the Danish high-risk subgroup was almost identical to unadjusted estimates from the UK cohort. Divergence in predictive ability between the cohorts was centred on the high-risk subgroup, which is based on the psychosocial subscale of the SBT. As seen in Study 1, which had examined 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).

Figure 3. The trajectory of recovery and care setting of intervention. Inspired by data from Pengel [65]




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 where 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 1c. Cohorts used for the comparison of the psychosocial profile across primary and secondary care

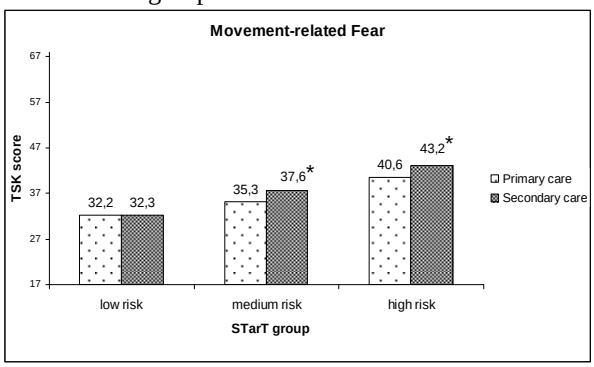
| Danish translation cohort from secondary care | Danish primary cohort (GP, PT) | Danish secondary predictive cohort | UK primary care development cohort | UK primary care BeBack cohort |
|---|---|---|---|---|
| Baseline only | Baseline only | | | |

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).

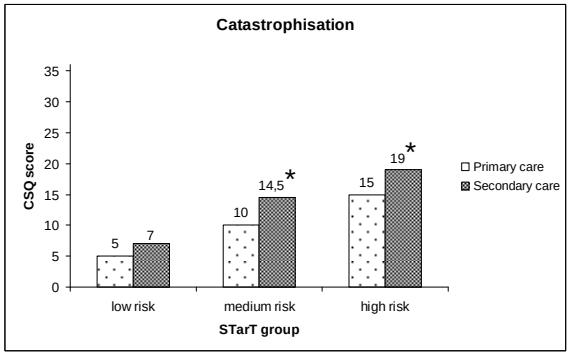
## Results

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).

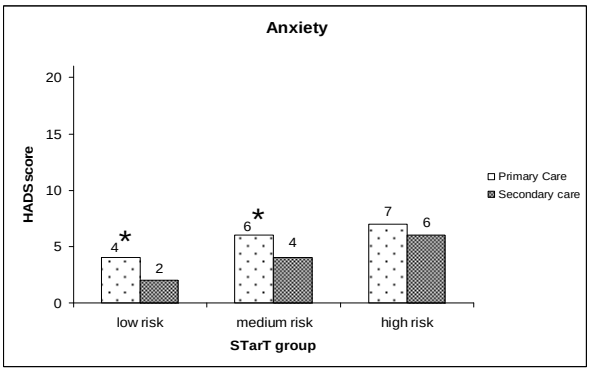
Figures 4. Differences between care settings on psychosocial construct scores, when stratified by SBT classification groups.



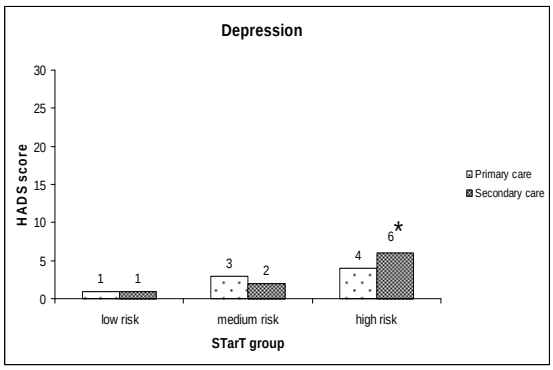

4a




4b




4c




4d

* 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

| Danish translation cohort from secondary care | Danish primary cohort (PT only) | Danish secondary predictive cohort | UK primary care development cohort | UK primary care BeBack cohort |
|---|---|---|---|---|
| | Baseline & 6-month outcomes | Baseline & 6-month outcomes | | |

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

cohorts and also calculated the relative risk of poor outcome using baseline pain intensity and baseline activity limitation as predictors.

## Results

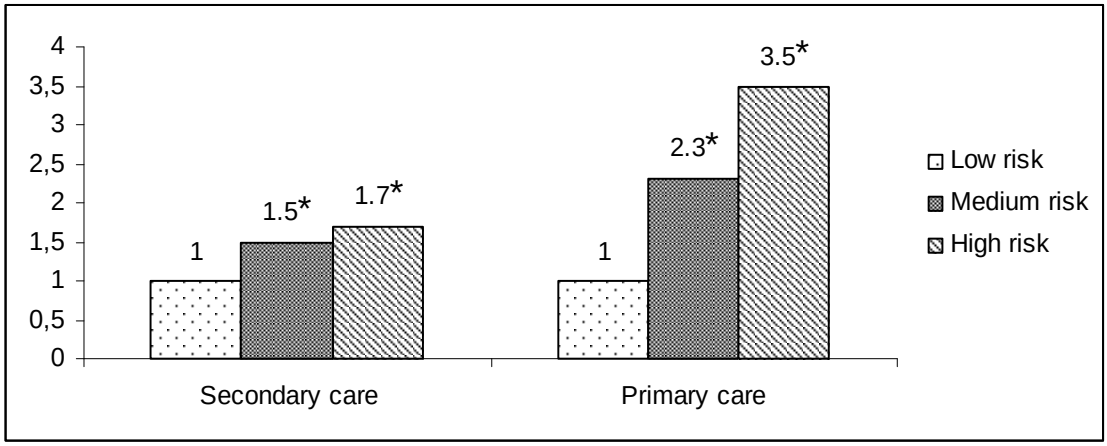
Overall, there were significant differences at baseline across the two cohorts on duration of episode and pain intensity. In concordance with earlier findings [33, 66], the pattern of increased score on pain intensity and activity limitation across the SBT subgroups from low risk to high risk was also 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.

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

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.

Figure 5. Relative risk of poor clinical outcome on activity limitation at 6-month follow-up by SBT subgroup in the Danish secondary and primary care cohorts.




*Significant differences

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*


(1) Walker BF, Muller R, Grant WD. Low back pain in Australian adults: prevalence and associated disability. J Manipulative Physiol Ther 2004 May;27(4):238-44.

(2) Leboeuf-Yde C, Nielsen J, Kyvik KO, Fejer R, Hartvigsen J. Pain in the lumbar, thoracic or cervical regions: do age and gender matter? A population-based study of 34,902 Danish twins 20-71 years of age. BMC Musculoskelet Disord 2009;10:39.

(3) Illemann Chrinstensen A, Ekholm O, Davidsen M, Juel K. Sundhed og sygelighed i Danmark 2010 & udviklingen siden 1987. The National Institute of Public Health, University of Southern Denmark; 2012.

(4) Bjerrum Koch M, Davidsen M, Juel K. De samfundsmæssige omkostninger ved rygsygdomme og rygsmerter i Danmark. The National Institute of Pulic Health, University of Southern Denmark; 2011.

(5) Hestbaek L, Leboeuf-Yde C, Manniche C. Is low back pain part of a general health pattern or is it a separate and distinctive entity? A critical literature review of comorbidity with low back pain. J Manipulative Physiol Ther 2003 May;26(4):243-52.

(6) Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012 Jun;64(6):2028-37.

(7) Delitto A. Research in low back pain: time to stop seeking the elusive "magic bullet". Phys Ther 2005 Mar;85(3):206-8.

(8) Deyo RA. Diagnostic evaluation of LBP: reaching a specific diagnosis is often impossible. Arch Intern Med 2002 Jul 8;162(13):1444-7.

(9) Fourney DR, Andersson G, Arnold PM, Dettori J, Cahana A, Fehlings MG, et al. Chronic low back pain: a heterogeneous condition with challenges for an evidence-based approach. Spine (Phila Pa 1976 ) 2011 Oct 1;36(21 Suppl):S1-S9.

(10) Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006 Mar;15 Suppl 2:S192-S300.

(11) Dagenais S, Tricco AC, Haldeman S. Synthesis of recommendations for the assessment and management of low back pain from recent clinical practice guidelines. Spine J 2010 Jun;10(6):514-29.

(12) Krismer M, van TM. Strategies for prevention and management of musculoskeletal conditions. Low back pain (non-specific). Best Pract Res Clin Rheumatol 2007 Feb;21(1):77-91.

(13) Laerum E, Storheim K, Brox JI. [New clinical guidelines for low back pain]. Tidsskr Nor Laegeforen 2007 Oct 18;127(20):2706.

(14) Liddle SD, Gracey JH, Baxter GD. Advice for the management of low back pain: a systematic review of randomised controlled trials. Man Ther 2007 Nov;12(4):310-27.

(15) Quebec Task Force. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. Spine (Phila Pa 1976 ) 1987 Sep;12(7 Suppl):S1-59.

(16) van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, et al. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006 Mar;15 Suppl 2:S169-S191.

(17) Dunn KM, Jordan KP, Croft PR. Contributions of prognostic factors for poor outcome in primary care low back pain patients. Eur J Pain 2011 Mar;15(3):313-9.

(18) Hill JC, Fritz JM. Psychosocial influences on low back pain, disability, and response to treatment. Phys Ther 2011 May;91(5):712-21.

(19) Fairbank J, Gwilym SE, France JC, Daffner SD, Dettori J, Hermsmeyer J, et al. The role of classification of chronic low back pain. Spine (Phila Pa 1976 ) 2011 Oct 1;36(21 Suppl):S19-S42.

(20) Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin 2006 Oct;22(10):1911-20.

(21) Petersen T, Olsen S, Laslett M, Thorsen H, Manniche C, Ekdahl C, et al. Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain. Aust J Physiother 2004;50(2):85-94.

(22) O'Sullivan P. It's time for change with the management of non-specific chronic low back pain. Br J Sports Med 2012 Mar;46(4):224-7.

(23) Dunn KM, Croft PR. Classification of low back pain in primary care: using "bothersomeness" to identify the most severe cases. Spine (Phila Pa 1976 ) 2005 Aug 15;30(16):1887-92.

(24) Chou R, Qaseem A, Snow V, Casey D, Cross JT, Jr., Shekelle P, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007 Oct 2;147(7):478-91.

(25) Burton AK, Tillotson KM, Main CJ, Hollis S. Psychosocial predictors of outcome in acute and subchronic low back trouble. Spine (Phila Pa 1976 ) 1995 Mar 15;20(6):722-8.

(26) Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. Clin J Pain 2003 Mar;19(2):80-6.

(27) Main CJ, Foster N, Buchbinder R. How important are back pain beliefs and expectations for satisfactory recovery from back pain? Best Pract Res Clin Rheumatol 2010 Apr;24(2):205-17.

(28) Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000 Apr;85(3):317-32.

(29) Haggman S, Maher CG, Refshauge KM. Screening for symptoms of depression by physical therapists managing low back pain. Phys Ther 2004 Dec;84(12):1157-66.

(30) Hill JC, Vohora K, Dunn KM, Main CJ, Hay EM. Comparing the STarT back screening tool's subgroup allocation of individual patients with that of independent clinical experts. Clin J Pain 2010 Nov;26(9):783-7.

(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.

(32) Hill JC, Dunn KM, Main CJ, Hay EM. Subgrouping low back pain: a comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. Eur J Pain 2010 Jan;14(1):83-9.

(33) Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008 May 15;59(5):632-41.

(34) Hilfiker R, Bachmann LM, Heitz CA, Lorenz T, Joronen H, Klipstein A. Value of predictive instruments to determine persisting restriction of function in patients with subacute non-specific low back pain. Systematic review. Eur Spine J 2007 Nov;16(11):1755-75.

(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.

(36) Hay EM, Dunn KM, Hill JC, Lewis M, Mason EE, Konstantinou K, et al. A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. BMC Musculoskelet Disord 2008;9:58.

(37) Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet 2011 Oct 29;378(9802):1560-71.

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

(39) Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. PLoS Med 2013 Feb;10(2):e1001380.

(40) Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. BMJ 2013;346:e5595.

(41) Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. BMJ 2013;346:e5793.

(42) Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Med 2013 Feb;10(2):e1001381.

(43) Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. Breast Cancer Res Treat 1998;52(1-3):289-303.

(44) Altman DG. Systematic reviews of evaluations of prognostic variables. BMJ 2001 Jul 28;323(7306):224-8.

(45) Hayden JA, Dunn KM, van der Windt DA, Shaw WS. What is the prognosis of back pain? Best Pract Res Clin Rheumatol 2010 Apr;24(2):167-79.

(46) Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009 May 28;338:b605.

(47) Henrica C.W.de Vet, Caroline B.Terwee, Lidwine B.Mokkink, Dirk L.Knol. Validity. Measrement in medicine. First ed. New York: Cambridge Universiry Press; 2011. p. 150-201.

(48) Foster NE, Bishop A, Thomas E, Main C, Horne R, Weinman J, et al. Illness perceptions of low back pain patients in primary care: what are they, do they change and are they associated with outcome? Pain 2008 May;136(1-2):177-87.

(49) Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Ann Intern Med 2004 Dec 21;141(12):920-8.

(50) Cleland JA, Fritz JM, Brennan GP. Predictive validity of initial fear avoidance beliefs in patients with low back pain receiving physical therapy: is the FABQ a useful screening tool for identifying patients at risk for a poor recovery? Eur Spine J 2008 Jan;17(1):70-9.

(51) Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976 ) 2000 Dec 15;25(24):3186-91.

(52) Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 1998 Nov;51(11):913-23.

(53) Ware JE, Jr., Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. Int J Technol Assess Health Care 1995;11(3):525-51.

(54) Albert HB, Jensen AM, Dahl D, Rasmussen MN. [Criteria validation of the Roland Morris questionnaire. A Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica]. Ugeskr Laeger 2003 Apr 28;165(18):1875-80.

(55) Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976 ) 1983 Mar;8(2):141-4.

(56) Swinkels-Meewisse EJ, Swinkels RA, Verbeek AL, Vlaeyen JW, Oostendorp RA. Psychometric properties of the Tampa Scale for kinesiophobia and the fear-avoidance beliefs questionnaire in acute low back pain. Man Ther 2003 Feb;8(1):29-36.

(57) Vlaeyen JW, Kole-Snijders AM, Boeren RG, van EH. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain 1995 Sep;62(3):363-72.

(58) Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77.

(59) Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-70.

(60) Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. Pain 1983 Sep;17(1):33-44.

(61) Swartzman LC, Gwadry FG, Shapiro AP, Teasell RW. The factor structure of the Coping Strategies Questionnaire. Pain 1994 Jun;57(3):311-6.

(62) Kirkwood BRSJAC. Measurement error: assessment and implications. Essential Medical Statistics. 2nd edition 2003 ed. Oxford: Blackwell Science Ltd.; 1988. p. 429-46.

(63) Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med 1999 Mar 16;130(6):515-24.

(64) Costa LC, Maher CG, McAuley JH, Hancock MJ, Herbert RD, Refshauge KM, et al. Prognosis for patients with chronic low back pain: inception cohort study. BMJ 2009;339:b3829.

(65) Pengel LH, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. BMJ 2003 Aug 9;327(7410):323.

(66) 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.

(67) Beneciuk JM, Bishop MD, Fritz JM, Robinson ME, Asal NR, Nisenzon AN, et al. The STarT Back Screening Tool and Individual Psychological Measures: Evaluation of Prognostic Capabilities for Low Back Pain Clinical Outcomes in Outpatient Physical Therapy Settings. Phys Ther 2012 Nov 2.

(68) Hill JC, Hay EM. Invited commentary. Phys Ther 2011 May;91(5):733-4.

(69) Albert HB, Briggs AM, Kent P, Byrhagen A, Hansen C, Kjaergaard K. The prevalence of MRI-defined spinal pathoanatomies and their association with modic changes in individuals seeking care for low back pain. Eur Spine J 2011 Aug;20(8):1355-62.

(70) Grotle M, Vollestad NK, Brox JI. Clinical course and impact of fear-avoidance beliefs in low back pain: prospective cohort study of acute and chronic low back pain: II. Spine (Phila Pa 1976 ) 2006 Apr 20;31(9):1038-46.

(71) Angst F, Verra ML, Lehmann S, Aeschlimann A. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. BMC Med Res Methodol 2008;8:26.

(72) Woby SR, Roach NK, Urmston M, Watson PJ. Psychometric properties of the TSK-11: a shortened version of the Tampa Scale for Kinesiophobia. Pain 2005 Sep;117(1-2):137-44.

(73) Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? JAMA 2010 Apr 7;303(13):1295-302.

(74) Nicholas MK, Linton SJ, Watson PJ, Main CJ. Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys Ther 2011 May;91(5):737-53.

(75) Bruyere O, Demoulin M, Brereton C, Humblet F, Flynn D, Hill JC, et al. Translation validation of a new back pain screening questionnaire (the STarT Back Screening Tool) in French. Arch Public Health 2012;70(1):12.

(76) Gusi N, Del Pozo-Cruz B, Olivares PR, Hernandez-Mocholi M, Hill JC. The Spanish version of the "STarT Back Screening Tool" (SBST) in different subgroups. Aten Primaria 2011 Feb 4.

(77) Kongsted A, Johannesen E, Leboeuf-Yde C. Feasibility of the STarT back screening tool in chiropractic clinics: a cross-sectional study of patients with low back pain. Chiropr Man Therap 2011;19:10.

(78) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-45.

(79) Hayden JA, Cote P, Steenstra IA, Bombardier C. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. J Clin Epidemiol 2008 Jun;61(6):552-60.

(80) Dunn KM, Croft PR. The importance of symptom duration in determining prognosis. Pain 2006 Mar;121(1-2):126-32.

(81) Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. Eur Spine J 2003 Apr;12(2):149-65.

(82) Albert HB, Godskesen M, Korsholm L, Westergaard JG. Risk factors in developing pregnancy-related pelvic girdle pain. Acta Obstet Gynecol Scand 2006;85(5):539-44.

(83) Sowden G, Hill JC, Konstantinou K, Khanna M, Main CJ, Salmon P, et al. Targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). Fam Pract 2012 Feb;29(1):50-62.

(84) Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine (Phila Pa 1976 ) 2002 Dec 15;27(24):2835-43.

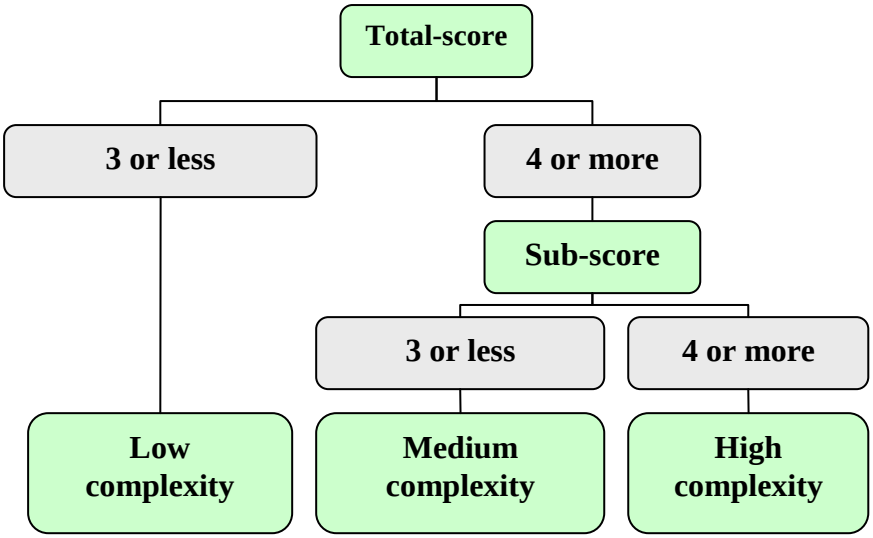
**PhD thesis - Appendices**

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

**STarT Spørgeskemaet**

Patientens navn: _____ Dato: _____

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

| | | **Nej** 0 | **Ja** 1 |
|---|---|---|---|
| 1 | I løbet af de seneste 2 uger har mine rygsmerter **bredt sig ned i mit/mine ben** | ☐ | ☐ |
| 2 | Jeg har haft smerter i mine **skuldre** eller **nakke** i løbet af de seneste 2 uger | ☐ | ☐ |
| 3 | Jeg har kun **gået korte afstande** på grund af mine rygsmerter | ☐ | ☐ |
| 4 | I løbet af de seneste 2 uger har jeg **klædt mig langsommere på** end normalt på grund af rygsmerter | ☐ | ☐ |
| | | **Uenig** 0 | **Enig** 1 |
| 5 | Det er egentligt ikke **sikkert** for en person i min tilstand at være fysisk aktiv | ☐ | ☐ |
| 6 | Jeg har været **bekymret** meget af tiden | ☐ | ☐ |
| 7 | Jeg føler mine rygsmerter er **forfærdelige** og de bliver **aldrig bedre** | ☐ | ☐ |
| 8 | Generelt har jeg ikke **nydt** alle de ting, som jeg plejede at nyde | ☐ | ☐ |

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

| Slet ikke | Lidt | Middel | Meget | Extremt |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |
| 0 | 0 | 0 | 1 | 1 |

**Total score (alle 9): _____** **Sub Score (spr. 5-9):_____**