

School of Physiotherapy and Exercise Science

**The Utility of the STarT Back Screening Tool in a Population with
Chronic Low Back Pain: A Prospective Study**

Michelle Kendell

**This thesis is presented for the degree of
Doctor of Clinical Physiotherapy
of
Curtin University**

February 2016

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

The research presented and reported in this thesis was conducted in accordance with the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research (2007) – updated March 2014. The proposed research study received human research ethics approval from the Curtin University Human Research Ethics Committee (EC00262), Approval Number PT262/2013.

Signature

Michelle Kendell

Date: 23 February 2016

Acknowledgements

I would first like to acknowledge and thank my supervisor Dr Darren Beales. Your generosity with your time and your guidance, input, and patience through this research project was greatly appreciated. Without you I would not have started on this research journey.

Professor Peter O'Sullivan, you have an amazing ability to see the big picture, translate research results into clinical relevance, and then tell the story. Thank you for your expert guidance in completing this research project.

Associate Professor Anne Smith, thank you for the time you took out of your busy schedule to run Journal Club, STATA Club, and share your vast knowledge of statistics. I am grateful for your assistance with the statistical analysis and your invaluable input into the project.

Mr Martin Rabey, thank you for sharing your data! I was the lucky recipient of your hard work to initially recruit and test a large sample of willing participants. I would also like to thank you for taking the time to answer my many questions and for your contribution to the final manuscript.

Dr Jonathan Hill, thank you for sharing your thoughts on the data analysis and on the results of this study.

Professor Tony Wright, thank you for guiding me through the Doctor of Clinical Physiotherapy degree. The path was rarely straight or clear. Without your time and effort, I would not have been able to navigate my way through the requirements of the degree.

Finally, I extend my sincere thanks and appreciation to my colleagues at the School of Physiotherapy and Exercise Science for your assistance and support over the last six years. Without this, I would not have been able to complete the Doctor of Clinical Physiotherapy degree.

Abstract

Study design

Cross sectional (Study 1) and prospective (Study 2).

Background

Chronic low back pain (LBP) is problematic with significant personal, social, and economic impact. The need to screen for indicators of poor prognosis and/or stratify patients with LBP based on risk is highlighted in the literature. The STarT Back Screening Tool (SBST) is one tool that has been developed to meet this need. There is a growing body of evidence for the psychometric properties of the SBST and for the tool's predictive and discriminative ability, particularly in populations with LBP of variable episode duration. Studies have shown that the SBST's performance has differed depending on the population in which it was evaluated. Clinical setting, cultural context, and LBP episode duration have all influenced the tool's performance. There are a lack of studies evaluating the utility of the SBST in a population exclusively with chronic LBP. While the SBST is commonly used in Australia, the tool is yet to be evaluated in an Australian context.

Aims

This study aimed to determine the utility of the SBST in a population exclusively with chronic non-specific LBP in Australia. The aim of the cross sectional study was to profile the SBST risk subgroups with respect to pain, disability, and psychological measures. The aim of the prospective study was to evaluate the predictive and discriminative ability of the SBST for pain intensity, self-reported LBP related disability, and global self-perceived change at one year follow-up.

Methods

Participants with dominant axial non-specific LBP of at least three months duration were recruited ($n=290$). At baseline participants completed the SBST and data were collected on demographic information, pain intensity (11-point numerical rating scale [NRS]), disability (Roland Morris Disability Questionnaire [RMDQ]), depression/anxiety/stress (Depression Anxiety Depression Scale), fear avoidance beliefs (Fear Avoidance Beliefs Questionnaire [FABQ]), catastrophising (Pain Catastrophising Scale), perceived risk of pain persistence (11-point NRS), self-efficacy (Pain Self-efficacy Questionnaire), and chronic pain acceptance

(Chronic Pain Acceptance Questionnaire). Baseline SBST risk subgroup differences were examined using a one-way analysis of variance (ANOVA), Kruskal-Wallis test, or Chi Squared test depending on the data type and distribution. Follow-up data were collected after one year ($n=264$). The follow-up measures were pain intensity (NRS), disability (RMDQ), and global perceived change (7-point global rating of change scale). The follow-up measures were dichotomised into recovered and not recovered for analysis. The proportion of participants who were not recovered with respect to each of the follow-up measure was calculated at a cohort level and by baseline SBST risk subgroup. The predictive ability of the SBST was evaluated with risk ratios (RR) using the low risk group as the reference category. In order to evaluate the discriminative ability of the SBST, receiver operating characteristic (ROC) curves were constructed and the area under the curve (AUC) was calculated for both the SBST total score and psychological subscale score for each of the follow-up measures. Likelihood ratios, sensitivity, specificity, and the diagnostic odds ratio (DOR) for the low risk group versus the medium/high risk group and the low/medium risk group versus the high risk group were calculated for the significant AUC values. Statistical significance was considered to be at the $p<0.05$ for all analyses.

Results

At baseline, the SBST categorised 82 participants (28.3%) as low risk, 116 (40.0%) as medium risk, and 92 (31.7%) as high risk. There were significant differences between the SBST risk subgroups for pain intensity ($p\leq 0.001$), disability ($p\leq 0.001$), and for the scores on the psychological questionnaires ($p\leq 0.001$ except FABQ work subscale $p=0.010$). The SBST risk subgroups demonstrated increasingly higher levels of pain, disability, and negative psychological affect and cognitions as subgroup categorisation increased from low to medium to high risk. Non-recovery rates at the one year follow-up were 76.1% ($n=201$) for pain, 31.4% ($n=83$) for disability, and 44.5% ($n=117$) for global perceived change. The predictive ability of the SBST was the strongest for disability. The medium risk group (RR 2.30 [95% confidence interval (CI) 1.28, 4.10], $p=0.003$) and the high risk group (RR 2.86 [1.60, 5.11], $p\leq 0.001$) had a meaningfully greater relative risk of being considered disabled compared to the low risk group. The medium risk group (RR 1.25 [1.04, 1.51], $p=0.013$) and the high risk group (RR 1.26 [1.03, 1.52], $p=0.020$) had a statistically greater risk of non-recovery with respect to pain but this relative risk was unlikely to be of any clinical significance. Although a higher proportion of both the medium and the high risk group perceived themselves as not improved at the one year follow-up compared to the low risk

group, the difference in risk was not significant (medium risk group RR 1.12 [0.78, 1.57], $p=0.558$; high risk group RR 1.26 [0.88, 1.79], $p=0.201$). The SBST's discriminative ability for pain was poor with an AUC [95% CI] of 0.63 [0.55, 0.71] for both the total score and the psychological subscale score. The SBST's discriminative ability for disability was acceptable with an AUC of 0.71 [0.64, 0.77] for the total score and 0.67 [0.60, 0.73] for the psychological subscale score. The SBST was unable to discriminate global perceived change with an AUC of 0.56 [0.49, 0.63] for the total score and 0.55 [0.48, 0.62] for the psychological subscale score. The positive and negative likelihood ratios for pain and disability were both relatively close to one. However, the positive likelihood ratios were higher and the negative likelihood ratios were lower for disability in comparison to pain, again suggesting the SBST performed better with respect to the disability measure.

Conclusions

The results of this study provide valuable information to clinicians on the usefulness and limitations of using the SBST in patients with chronic LBP in an Australian setting. The SBST should be applied with caution for patients presenting with chronic LBP. The SBST has value as a substitute for the administration of multiple full-length, unidimensional questionnaires to initially screen for high levels of pain, disability, and negative psychological affect and cognitions in clinical practice. Therefore the tool can alert the clinician to the need for further assessment. The SBST has moderate predictive and acceptable discriminative ability for future disability. However, the SBST's predictive and discriminative ability was relatively weak for future pain and the tool was unable to identify those participants who perceived themselves as improved versus not improved at the one year follow-up.

List of Abbreviations

ANOVA	Analysis of variance
AUC	Area under the curve
BQ	Bournemouth Questionnaire
CI	Confidence interval
CPAQ	Chronic Pain Acceptance Questionnaire
CSQ	Coping Strategies Questionnaire
DASS	Depression Anxiety Stress Scale
DOR	Diagnostic odds ratio
FABQ	Fear Avoidance Beliefs Questionnaire
FABQ-PA	FABQ – physical activity subscale
FABQ-W	FABQ – work subscale
GP	General Practice
GRCS	Global Rating of Change Scale
HADS	Hospital Anxiety Depression Scale
IQR	Interquartile range
LBP	Low back pain
LR-	Negative likelihood ratio
LR+	Positive likelihood ratio
MCID	Minimal clinically important difference
MDI	Major Depression Inventory
MIC	Minimal important change
NRS	Numerical Rating Scale
ODQ	Oswestry Disability Questionnaire
Örebro	Örebro Musculoskeletal Pain Screening Questionnaire

PCS	Pain Catastrophising Scale
PGIC	Patient Global Impression of Change
PHQ	Patient Health Questionnaire
PSEQ	Pain Self-efficacy Questionnaire
RCT	Randomised controlled trial
RMDQ	Roland Morris Disability Questionnaire
ROC	Receiver operating characteristic
RR	Risk ratio
SBST	STarT Back Screening Tool
<i>SD</i>	Standard deviation
T12	Twelfth thoracic vertebrae
TSK	Tampa Scale for Kinesiophobia
UK	United Kingdom
USA	United States of America

Table of Contents

Declaration.....	2
Acknowledgements.....	3
Abstract.....	4
List of Abbreviations	7
List of Figures	12
List of Tables	13
List of Appendices	14
Chapter 1: Introduction and Literature Review	15
1.1 Introduction.....	15
1.2 Prognostic factors in chronic low back pain	17
1.3 Screening for prognostic factors.....	18
1.3.1 Use of screening tools	18
1.3.2 Unidimensional versus multidimensional screening tools.....	20
1.4 STarT Back Screening Tool.....	21
1.4.1 Development and validation.....	21
1.4.2 Other studies evaluating the STarT Back Screening Tool.....	25
1.4.3 Studies in low back pain of variable duration	28
1.4.4 Studies in acute low back pain	31
1.4.5 Studies in chronic low back pain	33
1.4.6 Summary - STarT Back Screening Tool performance and low back pain episode duration.....	35
1.5 Limitations in the current research	35
1.6 Summary of key points	36
1.7 Basis of current research	36
1.7.1 Study 1.....	37
1.7.2 Study 2.....	37
1.8 References	38
Chapter 2: Cross Sectional Study	47
2.1 Aim.....	47
2.2 Methods.....	47
2.2.1 Participants.....	48
2.2.2 Sample size	49
2.2.3 Procedures	49

2.2.4 Demographic variables and clinical measures of pain and disability	49
2.2.5 STarT Back Screening Tool.....	50
2.2.6 Psychological measures.....	51
2.2.6.1 <i>Depression Anxiety Stress Scale</i>	51
2.2.6.2 <i>Fear Avoidance Beliefs Questionnaire</i>	51
2.2.6.3 <i>Pain Catastrophising Scale</i>	52
2.2.6.4 <i>Perceived risk of pain persistence</i>	52
2.2.6.5 <i>Pain Self-efficacy Questionnaire</i>	52
2.2.6.6 <i>Chronic Pain Acceptance Questionnaire</i>	52
2.2.7 Rationale for the chosen psychological measures	53
2.2.8 Data analysis.....	54
2.3 Ethics.....	54
2.4 Results.....	55
2.5 Key findings.....	63
2.6 References	64
Chapter 3: Prospective Study.....	68
3.0 Aim.....	68
3.1 Methods.....	68
3.1.1 Participants.....	68
3.1.2 Procedures	68
3.1.3 Follow-up measures	69
3.1.4 Data analysis.....	70
3.2 Ethics.....	72
3.3 Results.....	72
3.4 Key findings.....	81
3.5 References	82
Chapter 4: Discussion.....	84
4.1 Summary of key findings	84
4.2 Baseline sample characteristics.....	85
4.2.1 Recruitment considerations and baseline risk subgroup proportions.....	85
4.2.2 Baseline demographic and clinical characteristics.....	90
4.3 Cross sectional profile of the STarT Back Screening Tool risk subgroups.....	92
4.3.1 Clinical measures of pain and disability	93
4.3.2 Psychological measures.....	94
4.3.3 Summary - clinically meaningful differences between risk subgroups.....	97

4.3.4 Comparisons to other studies	97
4.3.5 STarT Back Screening Tool as a substitute for full-length, unidimensional measures	98
4.4 The one year follow-up sample characteristics.....	99
4.4.1 STarT Back Screening Tool risk subgroup stratification at follow-up.....	100
4.4.2 Non-recovery rates at follow-up	101
4.5 Predictive and discriminative ability of the STarT Back Screening Tool.....	103
4.5.1 Pain and disability	103
4.5.2 Global perceived change	108
4.5.3 Reasons for the relatively weak performance of the STarT Back Screening Tool in the current study.....	109
4.6 Strengths of the current study	111
4.7 Limitations of the current study	112
4.8 Implications for clinical practice	113
4.9 Recommendations for future research	114
4.10 Conclusions.....	116
4.11 References	117
Appendix A: Ethics Approval	124
Appendix B: Participant Information Sheet	125
Appendix C: Written Informed Consent Sheet	128
Appendix D: Discussion Supplement - Subgroup Proportions.....	129
Appendix E: Discussion Supplement - Follow-up Disability	131
Appendix F: Discussion Supplement - Disability Risk Ratios	133
Appendix G: Discussion Supplement - Disability Area Under the Curve	134
Appendix H: Discussion Supplement - Disability Likelihood Ratios	136

List of Figures

Figure 2.1	Flow of participants through the cross sectional study.
Figure 3.1	Flow of participants through the prospective study.
Figure 3.2	Relative risk of non-recovery at the one year follow-up by SBST risk subgroup.
Figure 3.3	ROC curves for the SBST baseline total score and psychological subscale score for pain ($\text{NRS} \geq 3$), disability ($\text{RMDQ} \geq 7$), and global perceived change ($\text{GRCS} \leq 0$) at the one year follow-up.

List of Tables

Table 1.1	Summary of the Predictive Utility of the STarT Back Screening Tool
Table 2.1	Descriptive Characteristics of Potential Participants Screened but not Eligible
Table 2.2	Descriptive Characteristics of Participants who did and did not Complete Baseline Assessment
Table 2.3	Baseline Descriptive Characteristics of Included Participants – Total Cohort and Stratified by STarT Back Screening Tool Risk Subgroup
Table 3.1	Baseline Descriptive Characteristics of Participants who did and did not Complete Follow-up
Table 3.2	Baseline vs. Follow-up STarT Back Screening Tool Risk Subgroup Stratification
Table 3.3	Descriptive Characteristics at One Year Follow-up – Total Cohort and Stratified by Baseline STarT Back Screening Tool Risk Subgroup
Table 3.4	Discriminative Ability of the STarT Back Screening Tool Baseline Total Score and Psychological Subscale Score using the AUC
Table 3.5	Discriminative Ability of the STarT Back Screening Tool Risk Subgroups - Likelihood Ratios, Sensitivity, Specificity, and the Diagnostic Odds Ratio
Table 4.1	Baseline Descriptive Characteristics across Key Studies

List of Appendices

Appendix A	Curtin University Human Research Ethics Committee Approval
Appendix B	Participant Information Sheet
Appendix C	Written Informed Consent Sheet
Appendix D	Discussion Supplement: Summary of Baseline STarT Back Screening Tool Risk Subgroup Proportions
Appendix E	Discussion Supplement: Proportion of Participants Disabled at Follow-up (RMDQ ≥ 7) – Total Cohort and Stratified by Baseline STarT Back Screening Tool Risk Subgroup
Appendix F	Discussion Supplement: Predictive ability of the STarT Back Screening Tool for Disability (RMDQ ≥ 7) using Risk Ratios
Appendix G	Discussion Supplement: Discriminative Ability of the STarT Back Screening Tool Baseline Total Score for Disability (RMDQ ≥ 7) using the AUC
Appendix H	Discussion Supplement: Discriminative Ability of the STarT Back Screening Tool Risk Subgroups for Disability (RMDQ ≥ 7) - Likelihood Ratios, Sensitivity, Specificity, and the Diagnostic Odds Ratio

Chapter 1: Introduction and Literature Review

1.1 Introduction

Chronic pain impacts upon a large proportion of the Australian population and is associated with interference of daily activities and high levels of psychological distress (Blyth et al., 2001). It has been reported that around 10-15% of patients with low back pain (LBP) develop chronic LBP (Balague et al., 2012). However, in a large prospective study in an Australian primary care setting, nearly one third of individuals with acute LBP did not recover from the presenting episode within one year (Henschke et al., 2008). Similarly, in a large prospective study in a primary care setting in the United States of America (USA), of the patients presenting with acute LBP, 25% were considered disabled and 26% had chronic pain two years later (Mehling et al., 2012). Worldwide, LBP is the leading cause of years lived with disability (Global Burden of Disease Study 2013 Collaborators, 2015). The economic burden of LBP in Australia has been estimated at \$9.17 billion per year (Walker et al., 2003). Healthcare practices require more effective and affordable strategies to better manage the rising burden of LBP (Vos et al., 2012).

Viewing chronic pain as a multidimensional biopsychosocial disorder is widely accepted (Gatchel et al., 2007, Waddell, 2006). Recognition of psychological factors such as anxiety, depression, catastrophising, low self-efficacy, and fear avoidance beliefs is considered to be essential to the effective management of chronic pain (Gatchel et al., 2007). Identifying and targeting modifiable risk factors for poor prognosis may be one strategy to improve the outcomes for patients with LBP (Foster et al., 2013, Foster, 2011, Hill et al., 2011). Clinical guidelines highlight the need to screen for indicators of poor prognosis in patients with both acute and chronic LBP (Accident Compensation Corporation, 2004, Airaksinen et al., 2006, Chou et al., 2007, van Tulder et al., 2006, Koes et al., 2010). One approach to identifying indicators of poor prognosis is through the use of screening tools (self-report questionnaires). A wide range of self-report screening tools to identify factors associated with poor prognosis and/or stratify patients based on risk have been developed (Accident Compensation Corporation, 2004, Du Bois et al., 2009, Fulton-Kehoe et al., 2008, Gabel et al., 2011, Gabel et al., 2013, Hazard et al., 1996, Hill et al., 2008, Linton and Boersma, 2003, Linton et al., 2011, Marhold et al., 2002, Neubauer et al., 2006, Reis et al., 2007, Sandborgh et al., 2007, Shaw et al., 2009a). The STarT Back Screening Tool (SBST) (Hill et al., 2008) is one such tool that screens for pain area, disability, and the presence of psychological factors. The SBST has received increasing interest in both clinical and research settings.

The tool has been advocated by The British Pain Society for early use (from two weeks after the onset of pain) in the management of low back and radicular pain (Lee et al., 2013). Stratified care that matches patients to treatment that targets their prognostic profile is a promising approach to improving outcomes for patients with LBP (Foster et al., 2013). A randomised controlled trial (RCT) demonstrated that stratified care based on the SBST risk subgroups resulted in better clinical outcomes and reduced costs compared to usual care in a United Kingdom (UK) primary care cohort with LBP of “mixed” or “variable” episode duration (i.e. participants reported acute, subacute, or chronic LBP) (Hill et al., 2011, Whitehurst et al., 2012).

Determining the validity and utility of commonly used screening measures to identify obstacles to recovery is a research priority (Main and George, 2011). The initial development and validation of the SBST was performed in a UK primary care setting where the participants had non-specific LBP of variable duration (Hill et al., 2008). In this original study by Hill and colleagues (2008), approximately 60% of both the developmental and independent external sample reported chronic LBP as defined by pain of at least three months duration (Chou et al., 2007, Gatchel et al., 2007, van Tulder et al., 2006). As prognostic factors may influence outcome differently depending on the clinical, social, and/or cultural context, investigation of a screening tool in different settings and populations is important in order to enhance the external validity and usefulness of the tool (Hockings et al., 2008, Hurley et al., 2001, Linton and Boersma, 2003, Linton and Hallden, 1998, Margison and French, 2007, Morsø et al., 2013b). Previous studies that have investigated the profile of the SBST risk subgroups and/or the prognostic value of the tool have clearly shown that the tool’s performance differed depending on the population in which it was evaluated (Field and Newell, 2012, Mehling et al., 2015, Morsø et al., 2013b, Morsø et al., 2014, Morsø et al., 2016, Newell et al., 2015). Low back pain episode duration has been reported to be associated with outcomes (Dunn and Croft, 2006) and has influenced the predictive performance of screening tools, including the SBST (Hill et al., 2008, Mehling et al., 2015, Morsø et al., 2016, Westman et al., 2008, Morsø et al., 2013b, Morsø et al., 2014). Authors have made clear recommendations that the utility of the SBST should be evaluated in different populations (Kongsted et al., 2011, Morsø et al., 2013b, Mehling et al., 2015) and that LBP episode duration is an important consideration (Morsø et al., 2016).

To date, only one small study has evaluated the performance of the SBST in a population exclusively with chronic LBP (Page et al., 2015). No previous studies have evaluated the predictive and discriminative ability of the SBST in an Australian context. Therefore the aims of this study were to: (i) profile the SBST risk subgroups with respect to clinical and psychological measures and (ii) investigate the predictive and discriminative ability of the SBST for pain intensity, self-reported disability, and global self-perceived change at one year follow-up in an Australian population exclusively with chronic non-specific LBP.

1.2 Prognostic factors in chronic low back pain

Prognostic factors for individuals with LBP are multidimensional and may include pain features, physical, psychological, social, work, lifestyle, and demographic factors.

Predictors of outcome are likely to be similar for acute and chronic LBP with psychological factors playing an important role (Balague et al., 2012). However, given that the natural course of acute and chronic LBP has been shown to be different (Dunn and Croft, 2006, Hayden et al., 2010, Grotle et al., 2010, Costa et al., 2012), it is likely that the specific factors involved, which combination of factors, or the strength of association of these factors with future outcomes are likely to differ depending on the duration of the LBP episode. Cultural differences are also likely to influence the association between prognostic factors and outcomes (Morsø et al., 2013b). In a large inception cohort of patients with LBP in an Australian primary care setting, there was an overlap in factors associated with a less favourable prognosis in patients initially presenting with acute LBP (Henschke et al., 2008) or chronic LBP (Costa et al., 2009). Overlapping factors included high baseline pain or disability and a high perceived risk of pain persistence. However, a number of factors associated with the outcomes differed between the patients presenting with acute versus chronic LBP in this cohort (Costa et al., 2009, Henschke et al., 2009). Comparably, Grotle and colleagues (2010) found that a similar set of prognostic indicators were associated with long term disability in both acute and chronic LBP in a primary care setting. These prognostic indicators included widespread pain, catastrophising, high baseline disability, being unemployed, and having a high Chronic Pain Grade (Grotle et al., 2010). However, only baseline disability explained a large proportion of the variance in 12 month disability outcome (Grotle et al., 2010).

A systematic review reported that there was little consistent evidence for which prognostic factors were associated with recovery from chronic LBP (Verkerk et al., 2012). Several

limitations related to the included studies in this systematic review were noted by the authors. These limitations were poor quality of methodology, differences in the measurement of the prognostic factors, differences in the measurement and reporting of the outcomes, and a lack of studies that included outcomes related to quality of life and global perceived effect (Verkerk et al., 2012). The relationship between prognostic factors and outcomes is complex and variable at a single person level whether considering acute or chronic LBP disorders. Regardless of these complexities, screening tools are a feasible way to initially screen for prognostic factors in a clinical setting.

1.3 Screening for prognostic factors

1.3.1 Use of screening tools

Screening for the risk of an unfavourable outcome in clinical practice is important. However, it has been reported that healthcare providers fail to routinely or systematically assess for risk factors, including psychological risk factors that may affect outcomes (Crawford et al., 2007, Daubs et al., 2010, Grevitt et al., 1998, Kent et al., 2009, Linton and Shaw, 2011, Main et al., 2010). Optimal methods of screening are also debated and may include an interview format, full-length unidimensional questionnaires, or brief or lengthy multidimensional tools.

Conflicting results are reported in the literature with respect to the value of using screening tools compared to using the clinician's impression, judgement, and intuition in order to identify risk factors for poor prognosis or individuals at risk of a poor outcome. Studies have reported that a clinician's ability to identify psychological factors such as depression (Haggman et al., 2004), psychological distress (Daubs et al., 2010, Grevitt et al., 1998), or fear avoidance beliefs (Calley et al., 2010) was poor in comparison to that of a screening tool. Similarly, other studies have reported that clinicians made inconsistent risk estimations using intuition alone in comparison to screening tools (Bishop and Foster, 2005, Hill et al., 2010a). Another study reported that clinicians were unable to accurately identify which patients were more likely to fail to improve over a course of treatment (Newell et al., 2013). Clinicians were also unable to discriminate between patients that were likely to have a poor pain or disability outcome (Kongsted et al., 2015). Unnecessary referrals based on the clinician's judgment rather than on formal screening may result in low risk patients receiving unnecessary treatment thus inflating costs or high risk patients undergoing less comprehensive treatment, resulting in poorer outcomes (Hill et al., 2011). In contrast,

other studies have reported that the clinician's prognostic assessment was similar to or even preferred over that of a screening tool (Dagfinrud et al., 2013, Jellema et al., 2007, Fersum et al., 2009, Kongsted et al., 2015, Karstens et al., 2015). A combination approach that includes the use of a screening questionnaire, clinical examination, and clinician judgement may be optimal to identify risk factors for poor prognosis or individuals at risk of a poor outcome (Beales et al., 2016).

There are several advantages to using a screening questionnaire as an adjunct to the clinical examination and clinician intuition and experience in identifying potential barriers to recovery. Clinicians in primary care settings are generally poorly equipped to assess psychological factors either due to time constraints to fully evaluate the multitude of variables in an interview format or due to insufficient training, knowledge, or skills (Crawford et al., 2007, Kent et al., 2009, Linton and Boersma, 2003, Overmeer et al., 2005). The use of screening questionnaires can enhance patient communication (Linton and Boersma, 2003, Nicholas et al., 2011) and may improve consistency in decision-making (Hill and Fritz, 2011).

There are limitations related to the use of screening questionnaires. The responder requires adequate language and literacy skills. Clinicians may feel screening tools are too lengthy and impractical to use routinely in a clinical setting (Mirkhil and Kent, 2009). Some clinicians may be reluctant or unprepared to explore non-medical domains (Shaw et al., 2009b). Others may not see the value in screening for psychological prognostic factors if they believe there is uncertainty about the effectiveness of interventions for patients with psychological barriers to recovery (Kent et al., 2009). No screening tool will be 100% accurate, therefore a patient's risk status as measured on a questionnaire may not reflect their true recovery trajectory. Screening also does not direct clinicians involved in a patient's care to a specific intervention (Hill and Fritz, 2011) nor can a screening tool make a sophisticated synthesis of patient preferences and expectations (Hill et al., 2010a). There may be a danger that access to treatment could be denied to individuals who have been misclassified as low risk on a screening tool alone. Finally, certain factors shown to demonstrate prognostic capabilities may not be modifiable through direct treatment such as age, sex, and pain episode duration. Despite the aforementioned limitations, it is generally accepted that the use of a brief screening tool to identify patients at risk of a poor outcome provides a basis for further assessment and for the development of targeted interventions most likely to facilitate recovery.

1.3.2 Unidimensional versus multidimensional screening tools

Many self-report screening questionnaires are unidimensional, consisting of several items each representing the same construct (Beneciuk et al., 2015). Others are multidimensional and quantify overall “risk complexity” or “cumulative prognostic risk” (Beneciuk et al., 2015, Hill et al., 2008, Wideman et al., 2012). A number of short, multidimensional tools have been developed to facilitate easy assessment of indicators of poor prognosis and/or stratify patients based on risk (Gabel et al., 2013, Hill et al., 2008, Linton et al., 2011).

There are several advantages to using concise, multidimensional screening tools over unidimensional ones. Individual prognostic factors have been shown to be relatively weak in predicting future outcomes (Cedraschi and Allaz, 2005, Chou and Shekelle, 2010). A combination of important psychological constructs may be a more useful indicator of patients at risk of a poor outcome (Campbell et al., 2013). Wideman and Sullivan (2012) demonstrated that the presence of a greater number of psychological risk factors (fear, depression, or catastrophising) reduced the likelihood of recovery with respect to pain and work disability. Similarly, measuring cumulative prognostic risk with the SBST total score was shown to be a better predictor of a broad range of outcomes compared to the tool’s single construct reference measures (Wideman et al., 2012). This literature supports that an increase in the number of risk factors present is linked to non-recovery and multidimensional tools can help identify this risk.

The use of a concise, multidimensional screening tool is more practical than administering a battery of unidimensional, full-length questionnaires in a clinical setting as it takes less time to complete and score, thus reducing responder and assessor burden. However, multidimensional tools do not provide detailed information on a specific construct which may mean that some patients require further assessment (Beneciuk et al., 2015, Beneciuk et al., 2013). Nonetheless, brief, multidimensional tools are useful for initial screening in order to guide decision-making as to whether or not further detailed assessment with single construct questionnaires is required (Hill and Fritz, 2011, Main and George, 2011). Multidimensional screening tools such as the SBST have the capacity to replace the use of numerous unidimensional tools as first-line screening (Beneciuk et al., 2015, Wideman et al., 2012) with further research required in different populations and settings to support this claim.

1.4 STarT Back Screening Tool

1.4.1 Development and validation

The SBST (<http://www.keele.ac.uk/sbst/startbacktool/>) (Keele University) is a brief 9-item, multidimensional screening tool used to identify patients with non-specific LBP who have potentially modifiable prognostic indicators for poor outcome, specifically future disability (Hill et al., 2008). It was primarily developed to subgroup patients into risk categories with cut-off scores based on baseline factors, rather than future outcome (Hill et al., 2010b).

The tool was designed to support primary care or first contact care decision-making. The development of the SBST was based on a review of the literature and a secondary analysis of a previously published RCT (Hay et al., 2005) and prospective study (Dunn and Croft, 2005) conducted in primary care settings in the UK. The constructs included in the tool were those shown to be the strongest independent predictors of future LBP related disability (Hill et al., 2008). The items in the SBST can be broadly categorised as either physical or psychological domains and cover eight constructs (referred leg pain, comorbid pain, disability [two items], fear avoidance, anxiety, catastrophising, depression, and back pain bothersomeness; the last five items represent the psychological subscale) (Hill et al., 2008). The question(s) related to each SBST construct were selected from previously validated questionnaires or validated single questions where available (Hill et al., 2008).

Items in the SBST use a dichotomised response format, *agree* (positive) or *disagree* (negative) with the exception of the question on back pain bothersomeness which is rated on a 5-point Likert scale (*not at all, slightly, moderately, very much, or extremely*) (Dunn and Croft, 2005). The response to this question is then dichotomised with a positive response equating to *very much* or *extremely* bothersome. The SBST is quick and easy to score with the overall total score produced by summing all positive responses with a possible overall total score of zero to nine with higher scores indicating greater risk complexity (Hill et al., 2008). The psychological subscale is independently summed for a score of zero to five (Hill et al., 2008). Based on cut-off scores, patients are allocated into three risk categories for targeted treatment, each directing a different care pathway. The low risk group has few negative prognostic indicators and is suitable for primary care management focusing on advice, analgesics, and education (Hill et al., 2008). The medium risk group has an unfavourable prognosis with high levels of physical prognostic indicators and is suited to physiotherapy intervention (Hill et al., 2008). The high risk group has a very unfavourable prognosis with a high level of psychological prognostic indicators and is suited to a

combination of cognitive-behavioural and physical management approaches (Hill et al., 2008).

The SBST's psychometric properties including discriminant validity, internal consistency, and test re-test reliability were assessed in a sample of participants with non-specific LBP in a UK primary care setting (i.e. the developmental sample) (Hill et al., 2008). Participants reported LBP of variable duration; approximately 60% of the cohort reported a LBP episode duration of greater than three months (Hill et al., 2008). Reference standards for the developmental sample included the Roland Morris Disability Questionnaire (RMDQ) for disability, the Pain Catastrophising Scale (PCS) for catastrophising, the Tampa Scale for Kinesiophobia (TSK) for fear, the Patient Health Questionnaire-2 (PHQ-2) for depression, the presence of referred leg pain, and back pain bothersomeness (Hill et al., 2008). The developmental sample was also used to identify cut-off scores with sensitivity and specificity values for allocation into the three risk groups (low, medium, and high) (Hill et al., 2008). In order to minimise the chance of subjecting patients to cognitive-behavioral management whom in fact did not require this approach, high specificity was given priority over high sensitivity when the optimal cut-off scores were determined (Hill et al., 2008). Study results showed acceptable to outstanding discriminant validity, acceptable levels of internal consistency, and acceptable test re-test reliability (Hill et al., 2008). Overall total scores of zero to three identified the low risk group, overall total scores of greater than three but less than four on the psychological subscale identified the medium risk group, and overall total scores of greater than three with four or more on the psychological subscale identified the high risk group (Hill et al., 2008). These cut-off scores resulted in a distribution of 39.7% ($n=52$) in the low risk group, 34.4% ($n=45$) in the medium risk group, and 25.2% ($n=33$) in the high risk group (Hill et al., 2008).

The external and predictive validity of the SBST was investigated in an independent external sample of participants with non-specific LBP in a primary care setting in the UK (Hill et al., 2008). Similar to the developmental sample, approximately 60% of the participants in the external sample reported LBP of greater than three months duration (Hill et al., 2008). Measurement and data collection procedures were the same as those for the developmental sample except that the PCS was replaced with the catastrophising subscale of the Coping Strategies Questionnaire (CSQ) and the PHQ-2 was replaced with the Hospital Anxiety Depression Scale (HADS). The use of different questionnaires from the original validation study ensured that the results were able to be generalised to the constructs

measured rather than the specific instrument (Hill et al., 2008). In the external sample, the proportion of participants in the low, medium, and high risk subgroup was 46.8% ($n=234$), 37.2% (186), and 14.8% ($n=74$) respectively which was similar to the developmental sample. At the six month follow-up, 16.7% ($n=39$) of the low risk group, 53.2% ($n=99$) of the medium risk group, and 78.4% ($n=58$) of the high risk group had a poor disability outcome as defined as greater than or equal to seven on the RMDQ. The low risk group versus the medium/high risk group had a positive likelihood ratio (LR+) of 2.32 [95% confidence interval (CI) 1.96, 2.76], negative likelihood ratio (LR-) of 0.30 [0.23, 0.40], sensitivity of 80.1%, and specificity of 65.4% for disability at the six month follow-up (Hill et al., 2008). The low/medium risk group versus the high risk group had a LR+ of 5.51 [3.30, 9.28], LR- of 0.74 [0.67, 0.81], sensitivity of 29.6% (corrected value reported as noted by Mehling and colleagues (2015)), and specificity of 94.6% for disability at the six month follow-up (Hill et al., 2008). In order to evaluate the influence of non-modifiable demographic factors on the SBST's performance, participants were stratified by age, sex, and pain episode duration (Hill et al., 2008). Age and sex had a small influence on the SBST's performance with the tool's discriminative ability better in older persons and men as demonstrated by higher LR+ and lower LR-. Low back pain episode duration had a stronger influence on the SBST's performance. The tool performed best when the duration of LBP ranged from one to six months (Hill et al., 2008).

Hill and colleagues (2010b) directly compared the SBST and Örebro Musculoskeletal Pain Screening Questionnaire (Örebro) in a UK primary care setting where participants reported LBP of variable duration. The Örebro is widely used in both clinical and research settings. The Örebro's reliability, validity, and predictive and discriminative ability has been extensively evaluated across multiple clinical settings, populations, and countries (Dunstan et al., 2005, Gabel et al., 2011, Grotle et al., 2007, Grotle et al., 2005, Grotle et al., 2006, Heneweer et al., 2007, Hockings et al., 2008, Hurley et al., 2001, Hurley et al., 2000, Jellema et al., 2007, Linton and Boersma, 2003, Linton and Hallden, 1998, Maher and Grotle, 2009, Margison and French, 2007, Vos et al., 2009, Westman et al., 2008). Completed SBST and Örebro data were available for 53% ($n=130$) of the participants, possibly introducing non-response bias to the results which would have had an equal effect on both tools (Hill et al., 2010b). In order to identify the low, medium, and high risk group using the SBST, cut-off scores as described by Hill and colleagues (2008) were used. A previously reported cut-off score of greater than or equal to 112 on the Örebro was used to identify patients at high risk while a score of 90 was used to separate the low from the medium risk group (Hurley et

al., 2001, Nordeman et al., 2006). Hill and colleagues (2010b) reported a high correlation between the SBST total score and the psychological subscale score with the Örebro score (0.80 and 0.77 respectively). The subgroup characteristics across the three risk groups were similar for both tools with the low risk groups almost identical in terms of pain and disability. The medium risk groups differed in that the SBST's medium risk group had slightly higher pain and disability scores but this was not found to be significant. The SBST allocated significantly fewer patients to the high risk group (SBST 25% and Örebro 38%) and significantly more patients to the medium risk group (SBST 35% and Örebro 22%) compared to the Örebro. The proportion of the sample in the low risk group was equal for both tools (40%). The discriminative ability of the SBST (total score and psychological subscale score) and Örebro were similar for disability, catastrophising, fear, co-morbid pain, time-off work, and episode duration when compared to reference standards (Hill et al., 2010b). The SBST was significantly better for discriminating bothersomeness and referred leg pain while the Örebro was significantly better for discriminating pain intensity (Hill et al., 2010b). This study supported the SBST's concurrent validity against a suitable reference standard (Hill et al., 2010b). The authors concluded that the SBST was an appropriate alternative for identifying high risk patients with LBP in a primary care setting (Hill et al., 2010b). It was not possible to compare the predictive ability of the SBST and Örebro due to the cross sectional nature of this study.

A limited number of studies have evaluated the SBST's responsiveness or sensitivity to change, an important psychometric property if a screening tool is to be considered useful as an outcome measure. Wideman and colleagues (2012) evaluated the SBST's sensitivity to physiotherapy treatment related change in comparison to the tool's corresponding full-length measures (RMDQ, PCS, TSK, and HADS-depression). The authors also evaluated the SBST's ability to detect clinically meaningful improvement in a number of outcomes at the four month follow-up (Wideman et al., 2012). The study was undertaken in a UK sample where participants reported LBP of variable duration (64% of the sample had chronic LBP) (Wideman et al., 2012). A reduction of three to five points on the SBST (corresponding to one risk category improvement) resulted in clinically meaningful improvements in perceived global change, pain, disability, pain catastrophising, fear of movement, and depression (Wideman et al., 2012). The authors concluded that the SBST could be used instead of its reference standards to imply changes in pain severity, disability, fear of movement, and catastrophising thus improving the efficiency of reassessment in a clinical setting (Wideman et al., 2012).

Other studies have provided additional preliminary support that a change in the SBST total score or a change in risk subgroup stratification may be useful to monitor the response to treatment for patients with LBP seeking physiotherapy care (Beneciuck et al., 2014, Beneciuk et al., 2013) and for office workers with subacute non-specific LBP undergoing an online occupational postural and exercise intervention (del Pozo-Cruz et al., 2012).

The SBST has been formally translated into several languages including Danish (Morsø et al., 2011), Spanish (Gusi et al., 2011), French (Bruyere et al., 2012), Brazilian Portuguese (Pilz et al., 2014), Mandarin (Luan et al., 2014), German (Aebischer et al., 2015), Persian (Abedi et al., 2015), and Finnish (Piironen et al., 2016). Several of these studies have also investigated the psychometric properties (e.g. discriminative ability, construct validity, internal consistency, and test re-test reliability) of the translated SBST with favourable results (Aebischer et al., 2015, Luan et al., 2014, Morsø et al., 2011, Pilz et al., 2014, Abedi et al., 2015, Piironen et al., 2016). To this investigator's knowledge, aside from the English version, only the predictive and discriminative ability of the Danish SBST has been evaluated (Kongsted et al., 2015, Morsø et al., 2013b, Morsø et al., 2014, Morsø et al., 2016).

1.4.2 Other studies evaluating the STarT Back Screening Tool

Since the original development and validation of the SBST in a primary care setting in the UK, a number of other studies have reported on various aspects of the performance of the SBST either using a cross sectional or prospective design. The tool has since been evaluated in the USA (Beneciuck et al., 2014, Beneciuk et al., 2013, Beneciuk et al., 2015, Fritz et al., 2011, George and Beneciuk, 2015, Mehling et al., 2015, Von Korff et al., 2014), Denmark (Kongsted et al., 2015, Kongsted et al., 2011, Morsø et al., 2013b, Morsø et al., 2013a, Morsø et al., 2014, Morsø et al., 2016), and Canada (Page et al., 2015). Several studies have specifically evaluated the SBST in a physiotherapy clinical setting (Beneciuck et al., 2014, Beneciuk et al., 2013, Beneciuk et al., 2015, Fritz et al., 2011, George and Beneciuk, 2015, Morsø et al., 2013b, Morsø et al., 2016) or chiropractic clinical setting (Field and Newell, 2012, Kongsted et al., 2015, Kongsted et al., 2011, Morsø et al., 2016, Newell et al., 2015, Page et al., 2015). Depending on the country, physiotherapy and chiropractic clinical settings have additionally been described as primary or secondary care by the authors. Outside of a physiotherapy or chiropractic clinical setting, other studies have evaluated the SBST in a general practice (GP) primary care cohort (Mehling et al., 2015, Morsø et al.,

2013b, Morsø et al., 2013a, Morsø et al., 2016, Von Korff et al., 2014, Wideman et al., 2012) or a cohort described as secondary care (Morsø et al., 2013a, Morsø et al., 2014, Morsø et al., 2016). The performance of the SBST has varied depending on the population in which the tool has been applied. A number of the key aforementioned studies will be discussed in further detail below. This discussion will be organised by LBP episode duration where chronic LBP is defined as pain present for a period of at least three months (Chou et al., 2007, Gatchel et al., 2007, van Tulder et al., 2006). A summary of the key studies specifically evaluating the predictive utility of the SBST is shown in Table 1.1.

Table 1.1. Summary of the Predictive Utility of the STarT Back Screening Tool

Study	Setting / population	Outcomes / follow-up	Predictive utility
Studies in LBP of variable duration			
Hill et al., 2008	Primary care UK ≈ 60% chronic LBP	Disability (RMDQ) / 6 months	Yes
Beneciuk et al., 2013	Physiotherapy USA 49% chronic LBP	Disability (ODQ), pain (NRS) / 6 months	Yes disability No pain
Beneciuck et al., 2014	Physiotherapy USA 49% chronic LBP	Disability (ODQ), pain (NRS) / 6 months	Yes disability No pain
George and Beneciuk, 2015	Physiotherapy USA 49% chronic LBP	Composite disability (RMDQ) & pain (NRS) / 6 months	Yes
Field and Newell, 2012	Chiropractic UK 31% chronic LBP	PGIC, pain, BQ / 14, 30, & 90 days	No
Newell et al., 2015	Chiropractic UK 47% chronic LBP	PGIC / 14, 30, & 90 days	No
Morsø et al., 2013b	Primary care Denmark (GP/Physiotherapy) ≈ 1/3 chronic LBP	Disability (RMDQ) / 3 months	Yes High risk group weaker vs. UK primary care
Studies in acute LBP			
Kongsted et al., 2015	Chiropractic Denmark 3% chronic LBP	Disability (RMDQ), pain (NRS) / 2 weeks; 3 & 12 months	Poor for disability No for pain
Mehling et al., 2015	Primary care USA 100% < 30 days LBP	Disability (RMDQ), Chronic Pain Risk, Composite of Perceived Recovery & Pain (NRS) / 6 months & 2 years	Poor for all outcomes
Studies in chronic LBP			
Morsø et al., 2014	Secondary care Denmark ≈ 80% chronic LBP	Disability (RMDQ), pain (NRS) / 6 months	Yes Weaker vs. Danish primary care
Page et al., 2015	Chiropractic Canada 100% chronic LBP	Disability (ODQ), pain (NRS), PGIC / 2, 4, 6, & 12 months	Yes disability Yes pain but weaker No for PGIC

LBP = low back pain; RMDQ = Roland Morris Disability Questionnaire; ODQ = Oswestry Disability Questionnaire; NRS = Numerical Rating Scale; PGIC = Patient Global Impression of Change; BQ = Bournemouth Questionnaire.

1.4.3 Studies in low back pain of variable duration

The majority of studies published since the original development and validation of the SBST have evaluated the tool in a population with LBP of mixed or variable episode duration (i.e. participants reported acute, subacute, or chronic LBP) (Beneciuk et al., 2014, Beneciuk et al., 2013, Beneciuk et al., 2015, Field and Newell, 2012, Fritz et al., 2011, George and Beneciuk, 2015, Kongsted et al., 2011, Morsø et al., 2013b, Newell et al., 2015, Von Korff et al., 2014, Wideman et al., 2012). Fritz and colleagues (2011) provided early preliminary support for the use of the SBST outside of a UK primary care setting. These authors examined the SBST in a physiotherapy outpatient setting in the USA in a sample where the participants reported a median LBP duration of 46 days (interquartile range [IQR] 18.5-147.0 days). The SBST risk subgroups were associated with baseline pain intensity and disability scores with increasingly higher scores on these clinical measures in the low to medium to high risk group, lending support to the construct validity of the tool (Fritz et al., 2011). This study did not evaluate the psychological profile of the SBST risk subgroups. However, a cross sectional study in a chiropractic clinical setting in Denmark (34% of the sample reported chronic LBP) in the same year reported a significant association between the SBST risk subgroups and scores on the Major Depression Inventory (MDI), Fear Avoidance Beliefs Questionnaire (FABQ), and the CSQ (catastrophising subscale) (Kongsted et al., 2011). Kongsted and colleagues (2011) concluded that the SBST risk subgroups were related to the presence of well-established psychological prognostic factors.

In addition to the study by Fritz and colleagues (2011), a number of other studies have investigated the SBST in a physiotherapy outpatient setting in the USA, both in a cross sectional and prospective manner. Beneciuk and colleagues (2013) reported that the SBST total score and the psychological subscale score were both significantly correlated with intake scores on a number of unidimensional psychological questionnaires (FABQ, PCS, TSK-11, and PHQ-9) with the strongest correlations obtained for the PCS, TSK-11, and PHQ-9. Prospectively, intake pain intensity (Numerical Rating Scale [NRS]) and disability (Oswestry Disability Questionnaire [ODQ]) scores were the strongest predictors of clinical outcomes at the six month follow-up with the SBST total score and the psychological subscale score adding to the prediction of disability but not pain (Beneciuk et al., 2013). Change scores on the SBST from intake to four week reassessment were also shown to be predictive of disability at the six month follow-up (Beneciuk et al., 2013). The authors suggested that a change in the SBST total score may be useful for treatment monitoring (Beneciuk et al., 2013).

Beneciuk and colleagues (2014) tested the predictive ability of the SBST risk subgroups using baseline categorisation, four week categorisation, and four week “change category” in the same physiotherapy outpatient setting in the USA. The authors’ rationale was that change scores/change categories may provide additional information for future outcomes beyond that achieved with initial intake screening only. Consistent with the results reported by Beneciuk and colleagues (2013), SBST intake categorisation (and four week change category) was not predictive of pain at the six month follow-up. The SBST baseline categorisation, four week categorisation, and four week change category were all predictive of disability at the six month follow-up with the predictive capabilities similar across all three variables (Beneciuk et al., 2014). The results of Beneciuk and colleagues (2013) and Beneciuk and colleagues (2014) together indicated that the SBST was not consistently predictive of future pain in a physiotherapy setting but the tool demonstrated predictive ability for future disability.

George and Beneciuk (2015) further evaluated the predictive ability of the SBST for a composite outcome measure of pain and disability in the same physiotherapy setting in the USA. The criteria for recovery was stringent with a pain intensity score of zero on a NRS combined with a RMDQ score of less than or equal to two equating to “recovered”. This was the first, and to date the only study to look at the ability of the SBST to predict outcome using a composite measure of recovery. The six month recovery rate was only 12.6% ($n=14$) using the aforementioned criteria which was completely driven by failure to meet the pain intensity score of zero according to the authors. The SBST was found to be a robust predictor of recovery status using this composite measure (George and Beneciuk, 2015). Higher baseline pain intensity and depressed mood (as measured by the PHQ-9) also uniquely contributed to non-recovery at the six month follow-up (George and Beneciuk, 2015). Fear avoidance, kinesiophobia, and depressive symptoms co-concurred with non-recovery (George and Beneciuk, 2015).

In a cross sectional study, Beneciuk and colleagues (2015) compared unidimensional and multidimensional approaches to screening for psychological distress, maladaptive coping, and disability in the same physiotherapy setting in the USA (49% of the sample reported chronic LBP). There was a risk dependent relationship between the SBST subgroups and a number of full-length unidimensional psychological and disability measures (Beneciuk et al., 2015). The authors concluded that the SBST was a viable replacement for the

administration of multiple unidimensional questionnaires to identify psychological distress, maladaptive coping, and disability thus providing further support for the use of the SBST in a physiotherapy setting.

Two studies have evaluated the predictive ability of the SBST in a UK chiropractic clinical setting where the participants reported LBP of variable duration (Field and Newell, 2012, Newell et al., 2015). Field and Newell (2012) reported that the SBST risk subgroups had no prognostic ability for differentiating outcomes of pain, Patient Global Impression of Change (PGIC) rating, or Bournemouth Questionnaire (BQ) scores (a multidimensional questionnaire which includes subscales for pain, disability, anxiety, depression, fear avoidance behavior, and locus of control) at 14, 30, or 90 days following the initial chiropractic visit. Newell and colleagues (2015) compared the predictive value of SBST intake scores to SBST scores two days following the initial chiropractic visit to help determine whether or not the timing of stratification was important. STarT Back Screening Tool stratification at intake and two days later had no prognostic ability for PGIC at the 90 day follow-up time point (Newell et al., 2015). STarT Back Screening Tool stratification at intake and two days later also showed an inconsistent and low association with PGIC outcome at the 14 and 30 day follow-up time point (Newell et al., 2015). There may be several reasons to explain why the SBST's predictive performance was poor in a UK chiropractic clinical setting. Although clinicians were blind to the SBST results, chiropractic treatment may have effectively addressed the modifiable risk factors, either consciously or unconsciously, which would have reduced the predictive strength of the tool (Field and Newell, 2012). Outcome measures were collected much earlier than the six month follow-up period used in the original SBST study (and subsequent studies) and this may have influenced the tool's performance. The outcome measures (PGIC and BQ) were also unique. Field and Newell (2012) and Newell and colleagues (2015) used change scores rather than absolute scores to measure outcome. The predictive utility of a tool can be different for absolute scores compared to change scores as highlighted by Kent and colleagues (2015) in their letter to the editor. The original validation of the SBST used absolute scores (Hill et al., 2008). Newell and colleagues (2015) proposed that the low response rate particularly at the 90 day follow-up time point could have resulted in response bias. Finally, it may be that the SBST is limited in its predictive capacity in a chiropractic clinical setting compared to a primary care or physiotherapy setting.

Using a validated Danish language version of the SBST (Morsø et al., 2011), the tool's predictive ability has been investigated in Denmark in a primary care setting with participants recruited from GP and physiotherapy clinics (Morsø et al., 2013b). The SBST distinguished between the low and the medium risk subgroup with similar predictive ability for pain and disability outcome at the three month follow-up in comparison to the original validation study by Hill and colleagues (2008). However, the predictive ability of the high risk subgroup was not as strong in the Danish primary care setting. This may have been due to cross cultural differences in the influence of psychological factors on outcomes between the Danish and UK samples (Morsø et al., 2013b).

In considering the published studies that have investigated the performance of the SBST in populations with LBP of variable duration, taken as a whole, the SBST appears to be a useful predictor of future disability in primary care and physiotherapy clinical settings both in the UK and outside of the UK. The tool's ability to predict future pain is less consistent. The SBST has not performed as well in chiropractic clinical settings or when using a patient's global perceived change as the measure of outcome.

1.4.4 Studies in acute low back pain

More recently, studies have explored the value of using the SBST in a population with acute LBP (Table 1.1). Kongsted and colleagues (2015) investigated the accuracy with which the SBST could predict two week, three month, and one year pain (NRS) and disability (RMDQ) outcomes in a chiropractic clinical setting in Denmark. In this study, 97% of the participants reported a LBP episode duration of less than three months with the majority (75%) having reported pain for less than two weeks. The SBST failed to discriminate pain outcomes and had poor discriminative ability for disability at all follow-up time points (area under the curve [AUC] ranged from 0.59-0.61). Likelihood ratios for the low, medium, and high risk subgroup were also all close to one indicating poor tool performance. In addition to LBP episode duration, other factors may have contributed to the SBST's poor prognostic performance. As an example, the threshold to define non-recovery in this study was a demanding standard where a poor outcome was defined as a pain intensity of greater than zero on a NRS and disability was defined as a score of greater than two on the RMDQ. The SBST's poor performance in this study by Kongsted and colleagues (2015) was in keeping with how the tool performed in other studies in UK chiropractic clinical settings (Field and Newell, 2012, Newell et al., 2015). Some possible reasons to explain the tool's poor performance in chiropractic clinical settings have been discussed above.

Mehling and colleagues (2015) and Von Korff and colleagues (2014) both evaluated the performance of an item set analogous to the existing SBST in a primary care setting in the USA. Mehling and colleagues (2015) evaluated this modified SBST item set in a cohort of patients with strictly defined acute LBP (less than 30 days duration) using the same cut-off scores as recommended for the original SBST. The modified SBST item set had limited ability to identify patients who developed chronic pain, disability, or those patients who perceived themselves as not recovered at both the six month and the two year follow-up with the AUC for the tool's total score ranging from 0.53-0.63 (Mehling et al., 2015). The positive predictive values were lower and the likelihood ratios were closer to one than those reported by Hill and colleagues (2008) for the disability outcome (greater than or equal to seven on the RMDQ for both studies). In contrast to the results reported by Mehling and colleagues (2015), this modified SBST item set performed better in a cohort of patients who reported LBP of variable duration (18% of the sample reported chronic LBP) (Von Korff et al., 2014). Von Korff and colleagues (2014) reported acceptable prediction of an unfavourable outcome at the four month follow-up as measured by the Graded Chronic Pain Scale (a combination measure of pain intensity and activity limitation) with an AUC of 0.75 [95% CI 0.70, 0.79]. In considering the results from the two aforementioned studies, LBP episode duration may at least partially account for the poor performance of this modified SBST item set. Mehling and colleagues (2015) concluded that the SBST should be used with caution in a patient group with acute LBP using existing cut-off scores but did acknowledge the need for a definitive study using the original SBST item set.

Morsø and colleagues (2016) performed a secondary analysis of data from three primary care cohorts and one secondary care cohort in order to evaluate the influence of care setting (chiropractic, physiotherapy, GP, and secondary care spine centre), LBP episode duration, and timing of follow-up on the prognostic ability of the SBST in Denmark. When using disability as the outcome measure, Morsø and colleagues (2016) reported that participant age, type of care setting, LBP episode duration, follow-up timing, and SBST risk subgroup were all associated with outcome. However, only episode duration was shown to affect the prognostic ability of the SBST risk subgroups (Morsø et al., 2016). The prognostic ability of the high risk group compared to the low risk group was stronger when the episode duration was greater than two weeks (Morsø et al., 2016). The prognostic ability of the medium risk group compared to the low risk group was also better when the episode duration was greater than 12 weeks compared to less than two weeks (Morsø et al., 2016).

Consistent with the results from other studies undertaken in populations where participants had a short duration of LBP, Morsø and colleagues (2016) concluded that the SBST's predictive ability was weak in very acute patients and that episode duration should be considered when using the SBST.

In summary, studies have shown that the SBST has limited prognostic ability in populations with acute or short duration LBP in comparison to populations with LBP of a variable or mixed duration. This is not surprising given that the SBST was not developed or initially validated in a population exclusively with acute LBP. In the original study by Hill and colleagues (2008), only 25% of the developmental sample and 17% of the external sample had pain of less than one month duration. Hill and colleagues (2008) also highlighted that LBP episode duration influenced the SBST's predictive performance. The tool performed best when the duration of LBP ranged from one to six months (Hill et al., 2008).

1.4.5 Studies in chronic low back pain

The majority of studies undertaken in a population with LBP of variable duration have recruited samples where at least one third of that sample reported LBP of greater than three months duration (Table 1.1). To date, only two studies have investigated the performance of the SBST in a population predominantly or exclusively with chronic LBP. The first study was in a secondary care setting in Denmark where approximately 80% of the participants reported chronic LBP (Morsø et al., 2014). The second study was a small study in Canada in a chiropractic clinical setting that investigated the SBST in a population exclusively with chronic LBP (Page et al., 2015).

The predictive and discriminative ability of the SBST for disability and pain outcome at six month follow-up was investigated by Morsø and colleagues (2014) and the results compared to those from an earlier study in a Danish primary care setting (Morsø et al., 2013b) which has been described previously in section 1.4.3. The predictive ability of the SBST for the disability outcome in the secondary care setting was not as strong as in the primary care setting even after adjusting for baseline differences in pain intensity and duration (Morsø et al., 2014). However, the SBST showed similar discriminative ability in both cohorts for pain and disability with the AUC ranging from 0.66-0.73. The weaker predictive ability reported in the Danish secondary care setting may have been related to different recovery trajectories of the two cohorts. The two cohorts had a similar level of disability at baseline, both at a cohort level and when stratified by the SBST, however, at

the six month follow-up, the secondary care cohort had a less favourable recovery course which may have attenuated the predictive strength of the SBST in this group (Morsø et al., 2014).

The study by Page and colleagues (2015) was the first and until now, the only study to investigate the SBST in a population exclusively with chronic LBP. The authors reported that the SBST total score was correlated with baseline disability (ODQ) and fear of movement (TSK) but not with baseline pain (NRS). The SBST total score correlated with the disability outcome at the two, four, six, and 12 month follow-up but only correlated with the pain outcome at the 12 month follow-up. In keeping with other studies that have used a PGIC scale as an outcome measure (Field and Newell, 2012, Newell et al., 2015), the SBST total score was not associated with PGIC at any of the follow-up time points. The SBST total score showed an excellent ability to discriminate disability at the six and the 12 month follow-up (AUC ranged from 0.84-0.82) and an acceptable ability to discriminate pain at the six and the 12 month follow-up (AUC ranged from 0.73-0.71). The SBST was unable to discriminate between those participants who had and had not improved according to the PGIC. Limitations of this study include the small sample size ($n=53$ at baseline) with a 13% loss at the six month follow-up and an 11% loss at the 12 month follow-up. The proportion of the sample with a poor outcome at each follow-up collection point was not reported. This study did not evaluate the association between SBST risk subgroups and outcomes or the psychological subscale score and outcomes.

In summary, there has been limited investigation of the SBST's performance in a population exclusively with chronic LBP. Furthermore, the aforementioned studies provide inconsistent results on the performance of the SBST in a group with chronic LBP. Although the SBST total score had an excellent ability to discriminate future disability in the study by Page and colleagues (2015), this study was small. In contrast, the SBST's discriminative ability for future disability was relatively weak in the study by Morsø and colleagues (2014). Morsø and colleagues (2014) also reported that the predictive ability of the SBST was relatively weak in comparison to that shown in a Danish primary care setting where the participants reported LBP of variable duration. The use of the SBST in a population with chronic LBP warrants further investigation.

1.4.6 Summary - STarT Back Screening Tool performance and low back pain episode duration

The literature has consistently demonstrated that the SBST risk subgroups are related to clinical and psychological measures when evaluated in a cross sectional manner. However, the predictive and discriminative strength of the SBST has shown variability across studies. There are several possible reasons as to why the tool's performance has differed across studies. These possibilities include: (a) differences in LBP episode duration of the samples, (b) differences in baseline characteristics of the samples including baseline risk stratification, (c) differences in natural recovery trajectories, (d) differences in the outcomes that were measured, (e) differences in how the outcomes were measured (e.g. absolute versus change scores; continuous vs dichotomised outcomes), (f) differences in the definitions/cut-off scores used to indicate non-recovery, (g) the timing of follow-up, (h) the clinical setting, (i) cultural influences, (j) the potential influence of treatment across the study period, (k) the statistical approaches used, and (l) differences in an individual's prognosis due to factors unrelated to the SBST. It should be noted that in the referenced studies where the cohort underwent some form of treatment (e.g. physiotherapy, chiropractic), this was not guided by the SBST risk subgroup or score but rather was at the discretion of the clinician. Clinicians and researchers cannot assume that the SBST will perform similarly across settings and populations. If the SBST is to be used with confidence in Australia for patients presenting with chronic LBP, the tool's performance must be evaluated in this group.

1.5 Limitations in the current research

There are a number of limitations in the current research that has evaluated the SBST. There is a lack of studies evaluating long term outcomes with only three original studies evaluating outcomes beyond a follow-up period of six months (Kongsted et al., 2015, Mehling et al., 2015, Page et al., 2015). Although global rating of change is a recommended core outcome measure (Dworkin et al., 2005), only a few studies have included this measure (Field and Newell, 2012, Mehling et al., 2015, Newell et al., 2015, Page et al., 2015, Wideman et al., 2012). The SBST has not been adequately investigated in a population exclusively with chronic LBP as there has only been one small study evaluating the tool's performance in this important group (Page et al., 2015). To date, no studies have evaluated the relationship between the SBST and perceived risk of pain persistence, self-efficacy, and

chronic pain acceptance, all important constructs. Finally, no studies have evaluated how the SBST performs in an Australian context.

1.6 Summary of key points

- Chronic pain can be associated with widespread suffering and limitations in physical, emotional, social, and occupational functioning (Breivik et al., 2006) resulting in significant personal, social, and economic impact.
- The importance, strength, and interaction of prognostic factors in acute and chronic LBP is likely to differ depending on the duration of the LBP episode.
- Identification and targeting of modifiable physical and psychological risk factors for poor prognosis may improve the outcomes for patients presenting with LBP.
- Screening tools to identify risk factors for poor prognosis and/or stratify patients with LBP based on risk have been described in the literature. The SBST has received significant attention in both clinical and research settings in recent years.
- Screening tools can facilitate efficient patient triage and guide clinical decision-making.
- Brief, multidimensional screening tools, such as the SBST, can easily be implemented in clinical practice and have several advantages when compared to unidimensional tools.
- There is a growing body of evidence for the psychometric properties, predictive ability, discriminative ability, and responsiveness of the SBST, particularly in groups with LBP of variable duration.
- Studies have shown that the SBST's performance has varied across populations with clinical setting, cultural context, and LBP episode duration all potentially influencing the tool's performance.

1.7 Basis of current research

As described above, there are limitations in the current literature on the utility of the SBST that warrants further investigation. The SBST risk subgroups have not been profiled nor has the tool's predictive and discriminative ability for a range of important measures been adequately investigated in a population exclusively with chronic non-specific LBP yet clinicians apply the tool in this group. Although the SBST is commonly used in Australia, the tool is yet to be evaluated in an Australian context. Before this short, multidimensional screening tool can be used with confidence to guide decision-making in the population of

interest, further evaluation is required. Therefore this study aimed to answer the following research question: What is the utility of the SBST in a population with chronic non-specific LBP in Australia?

1.7.1 Study 1

The first study was a cross sectional design profiling the SBST risk subgroups with respect to clinical and psychological measures in an Australian population exclusively with chronic non-specific LBP. The aim of this study was to describe the distribution of demographic variables, clinical measures of pain and disability, and psychological measures in relation to the SBST risk subgroups. Specifically, Study 1 aimed to examine whether the three SBST risk subgroups differed with respect to: (a) demographic variables, (b) clinical measures of pain and disability, and (c) psychological profile. The results of this study could help inform the implementation of the SBST by providing valuable information as to whether the SBST would be an appropriate substitute for the administration of multiple full-length, unidimensional questionnaires in order to determine the risk profile of a patient presenting with chronic LBP in Australia.

1.7.2 Study 2

The second study was a prospective design evaluating the predictive and discriminative ability of the SBST for pain intensity, self-reported LBP related disability, and global self-perceived change at one year follow-up in an Australian population exclusively with chronic non-specific LBP.

1.8 References

- ABEDI, M., MANSHADI, F., KHALKHALI, M., MOUSAVI, S., BAGHBAN, A., MONTAZERI, A. & PARNIANPOUR, M. 2015. Translation and validation of the Persian version of the STarT Back Screening Tool in patients with nonspecific low back pain. *Manual Therapy*, 20, 850-854.
- ACCIDENT COMPENSATION CORPORATION. 2004. *ACC 1038 New Zealand acute low back pain guide: incorporating the guide to assessing psychosocial yellow flags in acute low back pain* [Online]. Wellington, New Zealand: Accident Compensation Corporation. Available: <http://www.acc.co.nz/search-results/index.htm?ssUserText=low+back+pain+guidelines> [Accessed 19 December 2015].
- AEBISCHER, B., HILL, J., HILFIKER, R. & KARSTENS, S. 2015. German translation and cross-cultural adaptation of the STarT Back Screening Tool. *Public Library of Science One*, 10, e0132068.
- AIRAKSINEN, O., BROX, J., CEDRASCHI, C., HILDEBRANDT, J., KLABER MOFFETT, J., KOVACS, F., MANNION, A., REIS, S., STAAL, J., URSIN, H. & ZANOLI, G. 2006. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. *European Spine Journal*, 15 (Supplement 2), S192-S300.
- BALAGUE, F., MANNION, A., PELLISE, F. & CEDRASCHI, C. 2012. Non-specific low back pain. *Lancet*, 379, 482-491.
- BEALES, D., KENDELL, M., CHANG, R., HÅMSØ, M., GREGORY, L., RICHARDSON, K. & O' SULLIVAN, P. 2016. Association between the 10 item Örebro Musculoskeletal Screening Questionnaire and physiotherapists' perception of the contribution of biopsychosocial factors in patients with musculoskeletal pain. *Manual Therapy*, 23, 48-55.
- BENECIUCK, J., FRITZ, J. & GEORGE, S. 2014. The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings. *Journal of Orthopaedic and Sports Physical Therapy*, 44, 656-664.
- BENECIUK, J., BISHOP, M., FRITZ, J., ROBINSON, M., ASAL, N., NISENZON, A. & GEORGE, S. 2013. The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Physical Therapy*, 93, 321-333.
- BENECIUK, J., ROBINSON, M. & GEORGE, S. 2015. Subgrouping for patients with low back pain: a multidimensional approach incorporating cluster analysis and the STarT Back Screening Tool. *The Journal of Pain*, 16, 19-30.
- BISHOP, A. & FOSTER, N. 2005. Do physical therapists in the United Kingdom recognize psychosocial factors in patients with acute low back pain? *Spine*, 30, 1316-1322.
- BLYTH, F., MARCH, L., BRNABIC, A., JORM, L., WILLIAMSON, M. & COUSINS, M. 2001. Chronic pain in Australia: a prevalence study. *Pain*, 89, 127-134.
- BREIVIK, H., COLLETT, B., VENTAFRIDDA, V., COHEN, R. & GALLACHER, D. 2006. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *European Journal of Pain*, 10, 287-333.

- BRUYERE, O., DEMOULIN, M., BRERETON, C., HUMBLET, F., FLYNN, D., HILL, J., MAQUET, D., VAN BEVEREN, J., REGINSTER, J., CRIELAARD, J. & DEMOULIN, C. 2012. Translation validation of a new back pain screening questionnaire (the STarT Back Screening Tool) in French. *Archives of Public Health*, 70, 12.
- CALLEY, D., JACKSON, S., COLLINS, H. & GEORGE, S. 2010. Identifying patient fear-avoidance beliefs by physical therapists managing patients with low back pain. *Journal of Orthopaedic and Sports Physical Therapy*, 40, 774-783.
- CAMPBELL, P., BISHOP, A., DUNN, K., MAIN, C., THOMAS, E. & FOSTER, N. 2013. Conceptual overlap of psychological constructs in low back pain. *Pain*, 154, 1783-1791.
- CEDRASCHI, C. & ALLAZ, A. 2005. How to identify patients with poor prognosis in daily clinical practice. *Best Practice and Research Clinical Rheumatology*, 19, 577-591.
- CHOU, R., QASEEM, A., SNOW, V., CASEY, D., CROSS, T., SHEKELLE, P. & OWENS, D. 2007. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Annals of Internal Medicine*, 147, 478-491.
- CHOU, R. & SHEKELLE, P. 2010. Will this patient develop persistent disabling low back pain? *Journal of the American Medical Association*, 303, 1295-1302.
- COSTA, L., MAHER, C., HANCOCK, M., MCAULEY, J., HERBERT, R. & COSTA, L. 2012. The prognosis of acute and persistent low-back pain: a meta-analysis. *Canadian Medical Association Journal*, 184, E613-E624.
- COSTA, L., MAHER, C., MCAULEY, J., HANCOCK, M., HERBERT, R., REFSHAUGE, K. & HENSCHKE, N. 2009. Prognosis for patients with chronic low back pain: inception cohort study. *British Medical Journal*, 339, b3829.
- CRAWFORD, C., RYAN, K. & SHIPTON, E. 2007. Exploring general practitioner identification and management of psychosocial yellow flags in acute low back pain. *New Zealand Medical Journal*, 120, U2536.
- DAGFINRUD, H., STORHEIM, K., MAGNUSSEN, L., ODEGAARD, T., HOFTANISKA, I., LARSEN, L., RINGSTAD, P., HATLEBREKKE, F. & GROTTLE, M. 2013. The predictive validity of the Örebro Musculoskeletal Pain Questionnaire and the clinicians' prognostic assessment following manual therapy treatment of patients with LBP and neck pain. *Manual Therapy*, 18, 124-129.
- DAUBS, M., PATEL, A., WILLICK, S., KENDALL, R., HANSEN, P., PETRON, D. & BRODKE, D. 2010. Clinical impression versus standardized questionnaire: the spinal surgeon's ability to assess psychological distress. *The Journal of Bone and Joint Surgery*, 92, 2878-2883.
- DEL POZO-CRUZ, B., PARRACA, J., DEL POZO-CRUZ, J., ADSUAR, J., HILL, J. & GUSI, N. 2012. An occupational, internet-based intervention to prevent chronicity in subacute lower back pain: a randomised controlled trial. *Journal of Rehabilitation Medicine* 44, 581-587.
- DU BOIS, M., SZPALSKI, M. & DONCEEL, P. 2009. Patients at risk of long-term sick leave because of low back pain. *The Spine Journal*, 9, 350-359.

- DUNN, K. & CROFT, P. 2005. Classification of low back pain in primary care: using "bothersomeness" to identify the most severe cases. *Spine*, 30, 1887-1892.
- DUNN, K. & CROFT, P. 2006. The importance of symptom duration in determining prognosis. *Pain*, 121, 126-132.
- DUNSTAN, D., COVIC, T., TYSON, G. & LENNIE, I. 2005. Does the Örebro Musculoskeletal Pain Questionnaire predict outcomes following a work-related compensable injury? *International Journal of Rehabilitation Research*, 28, 369-370.
- DWORKIN, R., TURK, D., FARRAR, J., HAYTHORNTHWAIT, J., JENSEN, M., KATZ, N., KERNS, R., STUCKI, G., ALLEN, R., BELLAMY, N., CARR, D., CHANDLER, J., COWAN, P., DIONNE, R., GALER, B., HERTZ, S., JADAD, A., KRAMER, L., MANNING, D., MARTIN, S., MCCORMICK, C., MCDERMOTT, M., MCGRATH, P., QUESSY, S., RAPPAPORT, B., ROBBINS, W., ROBINSON, J., ROTHMAN, M., ROYAL, M., SIMON, L., STAUFFER, J., STEIN, W., TOLLETT, J., WERNICKE, J. & WITTER, J. 2005. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, 113, 9-19.
- FERSUM, K., O'SULLIVAN, P., KVÅLE, A. & SKOUEN, J. 2009. Inter-examiner reliability of a classification system for patients with non-specific low back pain. *Manual Therapy*, 14, 555-561.
- FIELD, J. & NEWELL, D. 2012. Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropractic and Manual Therapies* 20, 17.
- FOSTER, N. 2011. Barriers and progress in the treatment of low back pain. *BMC Medicine*, 9, 108.
- FOSTER, N., HILL, J., O'SULLIVAN, P. & HANCOCK, M. 2013. Stratified models of care. *Best Practice and Research Clinical Rheumatology*, 27, 649-661.
- FRITZ, J., BENECIUK, J., GEORGE, S., HILL, J. & HAY, E. 2011. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Physical Therapy*, 91, 722-736.
- FULTON-KEHOE, D., STOVER, B., TURNER, J., SHEPPARD, L., GLUCK, J., WICKIZER, T. & FRANKLIN, G. 2008. Development of a brief questionnaire to predict long-term disability. *Journal of Environmental and Occupational Medicine*, 50, 1042-1052.
- GABEL, C., BURKETT, B. & MELLOH, M. 2013. The shortened Örebro Musculoskeletal Screening Questionnaire: evaluation in a work-injured population. *Manual Therapy*, 18, 378-385.
- GABEL, C., MELLOH, M., YELLAND, M., BURKETT, B. & ROIKO, A. 2011. Predictive ability of a modified Örebro Musculoskeletal Pain Questionnaire in an acute/subacute low back pain working population. *European Spine Journal*, 20, 449-457.
- GATCHEL, R., PENG, Y., PETERS, M., FUCHS, P. & TURK, D. 2007. The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological Bulletin*, 133, 581-624.
- GEORGE, S. & BENECIUK, J. 2015. Psychological predictors of recovery from low back pain: a prospective study. *BMC Musculoskeletal Disorders*, 16, 49.

- GLOBAL BURDEN OF DISEASE STUDY 2013 COLLABORATORS 2015. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 386, 743-800.
- GREVITT, M., PANDE, K., O'DOWD, J. & WEBB, J. 1998. Do first impressions count? A comparison of subjective and psychologic assessment of spinal patients. *European Spine Journal*, 7, 218-223.
- GROTLE, M., BROX, J., GLOMSROD, B., LONN, J. & VOLLESTAD, N. 2007. Prognostic factors in first-time care seekers due to acute low back pain. *European Journal of Pain*, 11, 290-298.
- GROTLE, M., BROX, J., VEIEROD, M., GLOMSROD, B., LONN, J. & VOLLESTAD, N. 2005. Clinical course and prognostic factors in acute low back pain. *Spine*, 30, 976-982.
- GROTLE, M., FOSTER, N., DUNN, K. & CROFT, P. 2010. Are prognostic indicators for poor outcome different for acute and chronic low back pain consulters in primary care? *Pain*, 151, 790-797.
- GROTLE, M., VOLLESTAD, N. & BROX, J. 2006. Screening for yellow flags in first-time acute low back pain: reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. *The Clinical Journal of Pain*, 22, 458-467.
- GUSI, N., DEL POZO-CRUZ, B., OLIVARES, P., HERNANDEZ-MOCHOLI, M. & HILL, J. 2011. The Spanish version of the "STarT Back Screening Tool" (SBST) in different subgroups. *Atencion Primaria*, 43, 356-361.
- HAGGMAN, S., MAHER, C. & REFSHAUGE, K. 2004. Screening for symptoms of depression by physical therapists managing low back pain. *Physical Therapy*, 84, 1157-1166.
- HAY, E., MULLIS, R., LEWIS, M., VOHORA, K., MAIN, C., WATSON, P., DZIEDZIC, K., SIM, J., MINNS LOWE, C. & CROFT, P. 2005. Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet*, 365, 2024-2030.
- HAYDEN, J., DUNN, K., VAN DER WINDT, D. & SHAW, W. 2010. What is the prognosis of back pain? *Best Practice and Research Clinical Rheumatology*, 24, 167-179.
- HAZARD, R., HAUGH, L., REID, S., PREBLE, J. & MACDONALD, L. 1996. Early prediction of chronic disability after occupational low back injury. *Spine*, 21, 945-951.
- HENEWEER, H., AUFDEMKAMPE, G., VAN TULDER, M., KIERS, H., STAPPAERTS, K. & VANHEES, L. 2007. Psychosocial variables in patients with (sub) acute low back pain. *Spine*, 32, 586-592.
- HENSCHKE, N., MAHER, C., REFSHAUGE, K., HERBERT, R., CUMMING, R., BLEASEL, J., YORK, J., DAS, A. & MCAULEY, J. 2008. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *British Medical Journal*, 337, a171.
- HENSCHKE, N., MAHER, C., REFSHAUGE, K., HERBERT, R., CUMMING, R., BLEASEL, J., YORK, J., DAS, A. & MCAULEY, J. 2009. Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. *Arthritis and Rheumatism*, 60, 3072-3080.

- HILL, J., DUNN, K., LEWIS, M., MULLIS, R., MAIN, C., FOSTER, N. & HAY, E. 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632-641.
- HILL, J., DUNN, K., MAIN, C. & HAY, E. 2010b. Subgrouping low back pain: a comparison of the STarT Back Tool with the Örebro Musculoskeletal Pain Screening Questionnaire. *European Journal of Pain*, 14, 83-89.
- HILL, J. & FRITZ, J. 2011. Psychosocial influences on low back pain, disability, and response to treatment. *Physical Therapy*, 91, 712-721.
- HILL, J., VOHORA, K., DUNN, K., MAIN, C. & HAY, E. 2010a. Comparing the STarT Back Screening Tool's subgroup allocation of individual patients with that of independent clinical experts. *The Clinical Journal of Pain*, 26, 783-787.
- HILL, J., WHITEHURST, D., LEWIS, M., BRYAN, S., DUNN, K., FOSTER, N., KONSTANTINOOU, K., MAIN, C., MASON, E., SOMERVILLE, S., SOWDEN, G., VOHORA, K. & HAY, E. 2011. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*, 378, 1560-1571.
- HOCKINGS, R., MCAULEY, J. & MAHER, C. 2008. A systematic review of the predictive ability of the Örebro Musculoskeletal Pain Questionnaire. *Spine*, 33, E494-E500.
- HURLEY, D., DUSOIR, T., MCDONOUGH, S., MOORE, A. & BAXTER, G. 2001. How effective is the Acute Low Back Pain Screening Questionnaire for predicting 1-year follow-up in patients with low back pain? *The Clinical Journal of Pain*, 17, 256-263.
- HURLEY, D., DUSOIR, T., MCDONOUGH, S., MOORE, A., LINTON, S. & BAXTER, G. 2000. Biopsychosocial screening questionnaire for patients with low back pain: preliminary report of utility in physiotherapy practice in Northern Ireland. *The Clinical Journal of Pain*, 16, 214-228.
- JELLEMA, P., VAN DER WINDT, D., VAN DER HORST, H., STALMAN, W. & BOUTER, L. 2007. Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. *British Journal of General Practice*, 57, 15-22.
- KARSTENS, S., JOOS, S., HILL, J., KRUG, K., SZECSENYI, J. & STEINHAUSER, J. 2015. General practitioners views of implementing a stratified treatment approach for low back pain in Germany: a qualitative study. *Public Library of Science One*, 10, e0136119.
- KEELE UNIVERSITY. *STarT Back* [Online]. Available: <http://www.keele.ac.uk/sbst/startbacktool/> [Accessed 21 October 2015].
- KENT, P., KEATING, J. & TAYLOR, N. 2009. Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. *Manual Therapy*, 14, 88-100.
- KENT, P., MORSØ, L. & KONGSTED, A. 2015. Letter to the editor concerning 'Using the STarT Back Tool: does timing of stratification matter?'. *Manual Therapy*, 20, e13.
- KOES, B., VAN TULDER, M., LIN, C., MACEDO, L., MCAULEY, J. & MAHER, C. 2010. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *European Spine Journal*, 19, 2075-2094.

- KONGSTED, A., ANDERSEN, C., HANSEN, M. & HESTBAEK, L. 2015. Prediction of outcome in patients with low back pain - a prospective cohort study comparing clinicians' predictions with those of the Start Back Tool. *Manual Therapy*, 21, 120-127.
- KONGSTED, A., JOHANNESSEN, E. & LEBOEUF-YDE, C. 2011. Feasibility of the STarT Back Screening Tool in chiropractic clinics: a cross-sectional study of patients with low back pain. *Chiropractic and Manual Therapies*, 19, 10.
- LEE, J., GUPTA, S., PRICE, C. & BARANOWSKI, A. 2013. Low back and radicular pain: a pathway for care developed by the British Pain Society. *British Journal of Anaesthesia*, 111, 112-120.
- LINTON, S. & BOERSMA, K. 2003. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. *The Clinical Journal of Pain*, 19, 80-86.
- LINTON, S. & HALLDEN, K. 1998. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *The Clinical Journal of Pain*, 14, 209-215.
- LINTON, S., NICHOLAS, M. & MACDONALD, S. 2011. Development of a short form of the Örebro Musculoskeletal Pain Screening Questionnaire. *Spine*, 36, 1891-1895.
- LINTON, S. & SHAW, W. 2011. Impact of psychological factors in the experience of pain. *Physical Therapy*, 91, 700-711.
- LUAN, S., MIN, Y., LI, G., LIN, C., LI, X., WU, S., MA, C. & HILL, J. 2014. Cross-cultural adaptation, reliability, and validity of the Chinese version of the STarT Back Screening Tool in patients with low back pain. *Spine*, 39, E974-E979.
- MAHER, C. & GROTE, M. 2009. Evaluation of the predictive validity of the Örebro Musculoskeletal Pain Screening Questionnaire. *The Clinical Journal of Pain*, 25, 666-670.
- MAIN, C., FOSTER, N. & BUCHBINDER, R. 2010. How important are back pain beliefs and expectations for satisfactory recovery from back pain? *Best Practice and Research Clinical Rheumatology*, 24, 205-217.
- MAIN, C. & GEORGE, S. 2011. Psychologically informed practice for management of low back pain: future directions in practice and research. *Physical Therapy*, 91, 820-824.
- MARGISON, D. & FRENCH, D. 2007. Predicting treatment failure in the subacute injury phase using the Örebro Musculoskeletal Pain Questionnaire: an observational prospective study in a workers' compensation system. *Journal of Occupational and Environmental Medicine*, 49, 59-67.
- MARHOLD, C., LINTON, S. & LENNART, M. 2002. Identification of obstacles for chronic pain patients to return to work: evaluation of a questionnaire. *Journal of Occupational Rehabilitation*, 12, 65-75.
- MEHLING, W., AVINS, A., ACREE, M., CAREY, T. & HECHT, F. 2015. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19, 439-446.

- MEHLING, W., GOPISETTY, V., BARTMESS, E., ACREE, M., PRESSMAN, A., GOLDBERG, H., HECHT, F., CAREY, T. & AVINS, A. 2012. The prognosis of acute low back pain in primary care in the United States: a 2-year prospective cohort study. *Spine*, 37, 678-684.
- MIRKHIL, S. & KENT, P. 2009. The diagnostic accuracy of brief screening questions for psychosocial risk factors of poor outcome from an episode of pain. *The Clinical Journal of Pain*, 25, 340-348.
- MORSØ, L., ALBERT, H., KENT, P., MANNICHE, C. & HILL, J. 2011. Translation and discriminative validation of the STarT Back Screening Tool into Danish. *European Spine Journal*, 20, 2166-2173.
- MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.
- MORSØ, L., KENT, P., ALBERT, H. & MANNICHE, C. 2013a. Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care? *Manual Therapy*, 18, 54-59.
- MORSØ, L., KENT, P., MANNICHE, C. & ALBERT, H. 2014. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine Journal*, 23, 120-128.
- MORSØ, L., KONGSTED, A., HESTBAEK, L. & KENT, P. 2016. The prognostic ability of the STarT Back Tool was affected by episode duration. *European Spine Journal*, 25, 936-944.
- NEUBAUER, E., JUNGE, A., PIRRON, P., SEEMANN, H. & SCHILTENWOLF, M. 2006. HKF-R 10 - Screening for predicting chronicity in acute low back pain (LBP): a prospective clinical trial. *European Journal of Pain*, 10, 559-566.
- NEWELL, D., FIELD, J. & POLLARD, D. 2015. Using the STarT Back Tool: does timing of stratification matter? *Manual Therapy*, 20, 533-539.
- NEWELL, D., FIELD, J. & VISNES, N. 2013. Prognostic accuracy of clinicians for back, neck and shoulder patients in routine practice. *Chiropractic and Manual Therapies*, 21, 42.
- NICHOLAS, M., LINTON, S., WATSON, P., MAIN, C. & "DECADE OF THE FLAGS" WORKING GROUP 2011. Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. *Physical Therapy*, 91, 737-753.
- NORDEMAN, L., NILSSON, B., MOLLER, M. & GUNNARSSON, R. 2006. Early access to physical therapy treatment for subacute low back pain in primary health care: a prospective randomized clinical trial. *The Clinical Journal of Pain*, 22, 505-511.
- OVERMEER, T., LINTON, S., HOLMQUIST, L., ERIKSSON, M. & ENGFELDT, P. 2005. Do evidence-based guidelines have an impact in primary care? A cross sectional study of Swedish physicians and physiotherapists. *Spine*, 30, 146-151.

- PAGE, I., ABBODD, J., O. SHAUGHNESSY, J., LAURENCELLE, L. & DESCARREAU, M. 2015. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskeletal Disorders*, 16, 201.
- PIIRONEN, S., PAANANEN, M., HAAPEA, M., HUPLI, M., ZITTING, P., RYYNANEN, K., TAKALA, E., KORNILOFF, K., HILL, J., HAKKINEN, A. & KARPPINEN, J. 2016. Transcultural adaption and psychometric properties of the STarT Back Screening Tool among Finnish low back pain patients. *European Spine Journal*, 25, 287-295.
- PILZ, B., VASCONCELOS, R., MARCONDES, F., LODOVICH, S., MELLO, W. & GROSSI, D. 2014. The Brazilian version of STarT Back Screening Tool - translation, cross-cultural adaptation and reliability. *Brazilian Journal of Physical Therapy*, 18, 453-461.
- REIS, S., BORKAN, J., VANRAALTE, R., TAMIR, A., DAHAN, R. & HERMONI, D. 2007. The LBP Patient Perception Scale: a new predictor of LBP episode outcomes among primary care patients. *Patient Education and Counseling*, 67, 191-195.
- SANDBORGH, M., LINDBERG, P. & DENISON, E. 2007. Pain Beliefs Screening Instrument: development and preliminary validation of a screening instrument for disabling persistent pain. *Journal of Rehabilitation Medicine*, 39, 461-466.
- SHAW, W., PRANSKY, G. & WINTERS, T. 2009a. The Back Disability Risk Questionnaire for work-related, acute low back pain: prediction of unresolved problems at 3-month follow-up. *Journal of Occupational and Environmental Medicine*, 51, 185-194.
- SHAW, W., VAN DER WINDT, D., MAIN, C., LOISEL, P. & LINTON, S. 2009b. Early patient screening and intervention to address individual-level occupational factors ("blue flags") in back disability. *Journal of Occupational Rehabilitation*, 19, 64-80.
- VAN TULDER, M., BECKER, A., BEKKERING, T., BREEN, A., GIL DEL REAL, M., HUTCHINSON, A., KOES, B., LAERUM, E. & MALMIVAARA, A. 2006. Chapter 3 European guidelines for the management of acute nonspecific low back pain in primary care. *European Spine Journal*, 15 (Supplement 2), S169-S191.
- VERKERK, K., LUIJSTERBURG, P., MIEDEMA, H., POOL-GOUDZWAARD, A. & KOES, B. 2012. Prognostic factors for recovery in chronic nonspecific low back pain: a systematic review. *Physical Therapy*, 92, 1093-1108.
- VON KORFF, M., SHORTREED, S., SAUNDERS, K., LERESCHE, L., BERLIN, J., STANG, P. & TURNER, J. 2014. Comparison of back pain prognostic risk stratification item sets. *The Journal of Pain*, 15, 81-89.
- VOS, C., VERHAGEN, A. & KOES, B. 2009. The ability of the Acute Low Back Pain Screening Questionnaire to predict sick leave in patients with acute neck pain. *Journal of Manipulative and Physiological Therapeutics*, 32, 178-183.
- VOS, T., FLAXMAN, A., NAGHAVI, M., LOZANO, R., MICHAUD, C., EZZATI, M., SHIBUYA, K., SALOMON, J., ABDALLA, S., ABOYANS, V., ABRAHAM, J., ACKERMAN, I., AGGARWAL, R., AHN, S., ALI, M., ALVARADO, M., ANDERSON, H., ANDERSON, L., ANDREWS, K. & ATKINSON, C. 2012. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380, 2163-2196.

- WADDELL, G. 2006. Preventing incapacity in people with musculoskeletal disorders. *British Medical Bulletin*, 77 and 78, 55-69.
- WALKER, B., MULLER, R. & GRANT, W. 2003. Low back pain in Australian adults: the economic burden. *Asia-Pacific Journal of Public Health*, 15, 79-87.
- WESTMAN, A., LINTON, S., OHRVIK, J., WAHLEN, P. & LEPPERT, J. 2008. Do psychosocial factors predict disability and health at a 3-year follow-up for patients with non-acute musculoskeletal pain? A validation of the Örebro Musculoskeletal Pain Screening Questionnaire. *European Journal of Pain*, 12, 641-649.
- WHITEHURST, D., BRYAN, S., LEWIS, M., HILL, J. & HAY, E. 2012. Exploring the cost-utility of stratified primary care management for low back pain compared with current best practice within risk-defined subgroups. *Annals of the Rheumatic Diseases*, 71, 1796-1802.
- WIDEMAN, T., HILL, J., MAIN, C., LEWIS, M., SULLIVAN, M. & HAY, E. 2012. Comparing the responsiveness of a brief, multidimensional risk screening tool for back pain to its unidimensional reference standards: the whole is greater than the sum of its parts. *Pain*, 153, 2182-2191.
- WIDEMAN, T. & SULLIVAN, M. 2012. Development of a cumulative psychosocial factor index for problematic recovery following work-related musculoskeletal injuries. *Physical Therapy*, 92, 58-68.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Chapter 2: Cross Sectional Study

2.1 Aim

This study was a cross sectional design profiling the STarT Back Screening Tool (SBST) risk subgroups with respect to clinical and psychological measures in an Australian population exclusively with chronic non-specific low back pain (LBP). The aim of this study was to describe the distribution of demographic variables, clinical measures of pain and disability, and psychological measures in relation to the SBST risk subgroups. Specifically, this study aimed to examine whether the three SBST risk subgroups differed with respect to: (a) demographic variables, (b) clinical measures of pain and disability, and (c) psychological profile.

It was hypothesised that: (i) there would be an incremental increase in pain intensity and disability in the SBST risk subgroups with the high risk subgroup having the highest level of pain and disability and (ii) the scores on the psychological measures would incrementally increase (or decrease for the positive constructs of self-efficacy and chronic pain acceptance) across the SBST risk subgroups with the high risk subgroup having the highest negative psychological affect and cognitions.

The results of this study could help inform the implementation of the SBST by providing valuable information as to whether the SBST would be an appropriate substitute for the administration of multiple full-length, unidimensional questionnaires in order to determine the risk profile of a patient presenting with chronic LBP in Australia.

2.2 Methods

A cross sectional analysis was performed on baseline data obtained from a longitudinal study which evaluated the presence of multidimensional subgroups in a population with chronic non-specific LBP (Rabey et al., 2015). Data were collected between November 15, 2012 and January 3, 2014. Participants were recruited from private physiotherapy, psychology, and pain management clinics in Perth, Western Australia and via multimedia advertisements (newspaper, social media, radio, and television) circulated throughout the general community in both metropolitan and regional Western Australia.

2.2.1 Participants

The inclusion criteria were:

- Age between 18 and 70 years inclusive.
- Dominant axial non-specific LBP. Non-specific LBP is defined as pain occurring primarily in the low back which cannot be attributed to specific pathology such as infection, malignancy, fracture, an inflammatory disorder, cauda equina syndrome, spinal stenosis, radiculopathy, or a radicular pain disorder (Balague et al., 2012, Chou et al., 2007). Dominant axial LBP was defined as pain between T12 and the gluteal fold with 60% or more LBP according to the following question: “Which situation describes your pain over the past four weeks the best? 100% of the pain in the low back, 80% of the pain in the low back and 20% in the leg(s), 60% of the pain in the low back and 40% in the leg(s), 50% of the pain in the low back and 50% in the leg(s), 40% of the pain in the low back and 60% in the leg(s), or 20% of the pain in the low back and 80% in the leg(s).” This question reliably differentiates patients with dominant LBP versus dominant leg pain (Wai et al., 2009), reducing the likelihood of participants with radiculopathy or radicular pain being included in this study.
- Low back pain present for greater than or equal to three months. Chronic pain is commonly defined as prolonged or persistent pain of at least three months duration (Chou et al., 2007, Gatchel et al., 2007, van Tulder et al., 2006).
- Baseline pain intensity in the last week of greater than or equal to two on an 11-point Numerical Rating Scale (NRS). Participants were asked: “On average over the past week how bad has your pain been?” with a scale anchor of zero indicating *no pain* and 10 indicating the *worst pain imaginable* (Dworkin et al., 2005, Manniche et al., 1994). The NRS is a valid, reliable, and responsive measure of pain intensity (Dworkin et al., 2005).
- Low back pain related disability of greater than or equal to five on the Roland Morris Disability Questionnaire (RMDQ). The RMDQ consists of 24 statements about activity limitations due to back pain. The response to each statement is marked as either *yes* or *no* where a positive (*yes*) response scores one point. Scores range from zero (no disability) to 24 (severely disabled). The RMDQ is valid, reliable, and responsive to change (Kuijer et al., 2005, Roland and Fairbank, 2000, Roland and Morris, 1983).

Participants were excluded if they were unable to understand English, were pregnant, had a diagnosed neurological condition, had serious spinal pathology (e.g. cancer, inflammatory arthropathy, or acute vertebral fracture), or had undergone spinal surgery within the previous six months (e.g. multi-level fusion, instrumentation, or discectomy). Participants were also excluded if they had bilateral dorsal wrist or hand pain as the previously mentioned longitudinal study from which this baseline data were obtained involved tissue sensitivity testing at the wrist (Rabey et al., 2015).

2.2.2 Sample size

The sample size for this study was predetermined by the aforementioned longitudinal study from which baseline data for the current study were obtained (Rabey et al., 2015). The number of participants recruited for this study was similar to previously published studies that have investigated the performance of the SBST (Beneciuk et al., 2015, Fritz et al., 2011, Morsø et al., 2013b).

2.2.3 Procedures

Potential participants completed an “inclusion criteria screening checklist” either in paper format or online (Qualtrics Research Suite software <https://curtin.asia.qualtrics.com/ControlPanel/>) depending on preference. Ambiguous responses were clarified by telephone. Eligible participants were subsequently sent a standardised self-report questionnaire pack which included questions related to demographic information, clinical measures (pain and disability), the SBST, and a battery of psychological questionnaires (described below). All questionnaires are freely available online. Participants attended a laboratory based physical testing session (data not used in this study) at the School of Physiotherapy and Exercise Science, Curtin University, Perth, Western Australia at which time the questionnaire pack was collected and checked for any obvious incomplete questions or questionnaires.

2.2.4 Demographic variables and clinical measures of pain and disability

Demographic information related to age, sex, education level, employment status, occupation, compensation status, pain duration, overall general health, and recruitment source into the study (general community or clinical setting) was obtained.

Participants were asked about their education level with the following question: “How many years have you spent in education (school, college, university, or professional education)?” (Dionne et al., 2001). With respect to employment status, participants were asked: “Are you currently in work (either paid or unpaid work e.g. student, housewife)?” Participants were also asked for their occupation and the responses were dichotomised into manual or sedentary occupations based upon the Australian and New Zealand Standard Classification of Occupations, Version 1.2 (Australian Bureau of Statistics, 2013). Participants who nominated their occupation as “retired”, “student”, or “housewife” were grouped together as “not working”. To determine compensation status, a single question (“I have a claim for compensation for my pain”) was taken from the Fear Avoidance Beliefs Questionnaire (FABQ) - work subscale (FABQ-W) (Waddell et al., 1993). Responses were then dichotomised as a positive or negative compensation status. Pain duration was elicited by asking participants: “How long have you had your back pain for?” with responses converted to months. Information on overall general health was collected using a reliable and valid single item from the COOP-WONCA charts (Van Weel et al., 2012). Participants rated their overall general health in the previous two weeks on a 5-point scale ranging from *excellent* (score of one) to *poor* (score of five). Baseline pain intensity in the previous week (NRS), RMDQ score, and percentage of LBP versus leg pain was retrieved from the inclusion criteria screening checklist.

2.2.5 STarT Back Screening Tool

The nine items in the SBST (<http://www.keele.ac.uk/sbst/startbacktool/>) (Keele University) can be broadly categorised as either physical or psychological domains and cover eight constructs (referred leg pain, comorbid pain, disability [two items], fear avoidance, anxiety, catastrophising, depression, and back pain bothersomeness; the last five items represent the psychological subscale) (Hill et al., 2008). Items in the SBST use a dichotomised response format, *agree* (positive) or *disagree* (negative) with the exception of the question on back pain bothersomeness which is rated on a 5-point Likert scale (*not at all, slightly, moderately, very much or extremely*) (Dunn and Croft, 2005). The response to this question is then dichotomised with a positive response equating to *very much* or *extremely* bothersome. The overall total score is produced by summing all positive responses with a possible overall total score of zero to nine with higher scores indicating greater risk complexity (Hill et al., 2008). The psychological subscale is independently summed for a score of zero to five (Hill et al., 2008). Based on cut-off scores, patients are allocated into three subgroups: low risk (total score less than or equal to three), medium risk (total score

greater than or equal to four with a psychological subscale score of less than or equal to three), or high risk (total score greater than or equal to four with a psychological subscale score of greater than or equal to four) (Hill et al., 2008).

The SBST's discriminant validity, internal consistency, test re-test reliability, external validity, and predictive validity has been reported by the original authors (Hill et al., 2008). Concurrent validity against a reference standard, the Örebro Musculoskeletal Pain Screening Questionnaire (Örebro), has also been established (Hill et al., 2010b). Further details on the development, validation, and psychometric properties of the SBST can be found in Chapter 1.

2.2.6 Psychological measures

2.2.6.1 Depression Anxiety Stress Scale

Depression, anxiety, and stress were measured with the 21-item short form of the Depression Anxiety Stress Scale (DASS-21) (Lovibond and Lovibond, 1995). The DASS-21 was selected as it evaluates all three constructs of psychological distress and the short form reduces responder burden. This questionnaire contains 21 quantitative statements (seven items in each subscale) each rated on a 4-point scale with zero indicating *did not apply to me at all*, one indicating *applied to me to some degree*, two indicating *applied to me to a considerable degree*, and three indicating *applied to me very much or most of the time*. Each subscale is summed and the score doubled (maximum score of 42 for each subscale) with higher scores representing higher levels of depression, anxiety, or stress. Based on the total score for each subscale, each construct can be classified as “normal”, “mild”, “moderate”, “severe”, or “extremely severe”. The reliability and validity of the DASS-21 has been established (Lovibond and Lovibond, 1995).

2.2.6.2 Fear Avoidance Beliefs Questionnaire

Fear avoidance beliefs were measured with the FABQ (Waddell et al., 1993). Statements are rated on a 7-point scale using the anchors *completely disagree* and *completely agree*. The questionnaire contains two subscales: (a) a 5-item physical activity subscale (FABQ-PA) where four items are scored; scores range from zero to 24 and (b) an 11-item work subscale (FABQ-W) where seven items are scored; scores range from zero to 42. Higher scores represent higher fear avoidance beliefs on both subscales. The reliability and validity of this questionnaire has been reported (George et al., 2010, Jacob et al., 2001, Swinkels-Meewisse et al., 2003, Waddell et al., 1993).

2.2.6.3 Pain Catastrophising Scale

Pain catastrophising was measured with the Pain Catastrophising Scale (PCS) (Sullivan et al., 1995). The PCS evaluates negative thoughts and feelings related to rumination, magnification, and helplessness (Sullivan et al., 1995). Rumination is described as worry and an inability to inhibit pain related thoughts (Sullivan et al., 1995). Magnification is an exaggeration of the unpleasantness of pain situations and expectancies for negative outcome (Sullivan et al., 1995). Helplessness is the inability to deal with painful situations (Sullivan et al., 1995). The PCS is the only instrument dedicated to the assessment of catastrophising and examines the three different components (rumination, magnification, and helplessness). Thirteen statements are rated on a 5-point scale ranging from *not at all* to *all the time* with the total score ranging from zero to 52. Higher scores indicate greater levels of pain catastrophising. The PCS's reliability and validity has been reported (Osman et al., 2000, Sullivan et al., 1995).

2.2.6.4 Perceived risk of pain persistence

Perceived risk of pain persistence was measured by a single self-report item taken from the Örebro (Linton and Boersma, 2003). Participants were asked: "In your view, how large is the risk that your current pain will become persistent?" (Linton and Boersma, 2003). A NRS is used, anchored at one end with zero indicating *no risk* and 10 at the other end indicating *a very large risk*.

2.2.6.5 Pain Self-efficacy Questionnaire

Pain self-efficacy was measured with the Pain Self-efficacy Questionnaire (PSEQ) (Nicholas, 2007). This questionnaire appears to be the most commonly used tool to assess this construct. The PSEQ measures confidence in performing 10 activities despite pain (Nicholas, 2007). Each statement is rated on a 7-point scale where zero equates to *not confident at all* and six equates to *completely confident*. Scores for each statement are summed with the total score ranging from zero to 60 with higher scores indicating stronger self-efficacy beliefs. The reliability and validity of the PSEQ has been established (Asghari and Nicholas, 2001, Kaivanto et al., 1995, Nicholas, 2007).

2.2.6.6 Chronic Pain Acceptance Questionnaire

Chronic pain acceptance was measured with the short form of the Chronic Pain Acceptance Questionnaire (CPAQ), the CPAQ-8 (Fish et al., 2010). The CPAQ is the most common tool

used to measure acceptance in chronic pain and its short form was selected to reduce responder burden. The CPAQ-8 consists of eight statements that are rated on a 7-point scale ranging from *never true* to *always true*. The questionnaire measures a person's ability to be in ongoing pain without trying to avoid it or control it and contains two subscales (pain willingness and activity engagement) each with four items. Each subscale is summed and has a score range of zero to 24. The total score on the CPAQ-8 ranges from zero to 48. Higher scores indicate higher levels of activity engagement and pain willingness (Fish et al., 2010, McCracken et al., 2004). The reliability and validity of the CPAQ has been reported (Fish et al., 2010).

2.2.7 Rationale for the chosen psychological measures

The constructs of depression, anxiety, fear avoidance, and catastrophising are core constructs found in the SBST with the DASS-21, FABQ, and PCS used as unidimensional reference standards in this study. The use of different questionnaires to measure these constructs compared to the SBST's original developmental and validation study ensured that the results were able to be generalised to the constructs measured rather than to the specific instrument (Hill et al., 2008). To date, no study has used the DASS or DASS-21 as a reference standard to evaluate the SBST.

Perceived risk of pain persistence, self-efficacy, and chronic pain acceptance are not core constructs found in the SBST. However, the question "I feel that my back pain is terrible and it's never going to get any better" may in part reflect a patient's perceived risk of pain persistence. To date, no study has investigated the relationship between the SBST risk subgroups and these psychological measures. Perceived risk of pain persistence has been linked to poorer outcomes in chronic LBP disorders (Costa et al., 2009, Foster et al., 2010, Laisne et al., 2012). Low self-efficacy has been shown to be predictive of disability outcomes in LBP (Foster et al., 2010). In a recent systematic review, higher levels of self-efficacy were reported to be associated with less functional impairment, less distress, and lower pain severity in chronic pain samples (Jackson et al., 2014). Greater pain acceptance has been shown to be associated with less distress and pain related disability in persistent chronic musculoskeletal pain disorders (McCracken and Eccleston, 2006, Wright et al., 2011). Greater pain acceptance has also been reported to be associated with better functioning at a future point in time in chronic musculoskeletal pain disorders, including chronic LBP (McCracken and Eccleston, 2005).

2.2.8 Data analysis

When calculating the total score of each questionnaire, any missing data were managed as suggested in the questionnaire's original manuscript where described. Otherwise, when one item was missing, the imputed average of the relevant scale or subscale was used in the calculation of the total score. If two or more items were missing, the questionnaire total was coded as missing. If answers to single question items were not given, they were coded as missing.

Descriptive statistics were calculated for the demographic variables, clinical measures of pain and disability, and all the psychological measures with respect to the total cohort and with respect to each SBST risk subgroup. Histograms and box plots were inspected for all continuous variables with normally distributed data expressed as means and standard deviations (*SD*). Skewed and nominal variables were expressed as medians and interquartile ranges (IQR). Frequency counts and percentages were used to express categorical variables.

STarT Back Screening Tool subgroup differences for the continuous demographic variables, clinical measures, and the psychological measures were examined using a one-way analysis of variance (ANOVA) for normally distributed variables and the Kruskal-Wallis test for variables with skewed data. STarT Back Screening Tool subgroup differences for categorical variables were examined using the Chi Squared test.

All data analyses were performed using Stata 13.1 (Statacorp, Texas, USA) for Windows. Statistical significance was considered to be $p < 0.05$ for all analyses.

2.3 Ethics

This study was approved by the Curtin University Human Research Ethics Committee (Approval Number PT262/2013) (Appendix A). The longitudinal study from which this study obtained baseline data (Rabey et al., 2015) was also approved by the Curtin University Human Research Ethics Committee (Approval Number HR112/2012). All participants received a Participant Information Sheet in plain language outlining the purpose of the study, methods, risks and discomforts, and the benefits of the research (Appendix B). Confidentiality was guaranteed and participants were made aware that they were able to

withdraw at any time without prejudice. Written informed consent was obtained from each participant (Appendix C).

2.4 Results

The flow of participants through the study is shown in Figure 2.1. During the study period, 586 volunteers were screened for study eligibility. Of these, 237 were excluded because they did not meet the inclusion criteria with the most common reason being a score of less than five on the RMDQ ($n=130$ [54.9%]) (Table 2.1). Four participants were excluded from data analysis as they were recruited from local public hospitals for which this study did not have ethics approval (Figure 2.1). Fifty five participants did not complete baseline data collection and were excluded from the study (Table 2.2). A comparison of the characteristics of the participants who did and did not complete baseline data collection indicated that there were no significant differences between these groups on age ($p=0.326$), sex ($p=0.566$), pain intensity ($p=0.282$), disability ($p=0.149$), or percentage of LBP versus leg pain ($p=0.667$) (Table 2.2). A total of 290 participants were eligible for inclusion in this study with 228 (78.6%) participants recruited from the general community and 62 (21.4%) from a clinical setting. Of the participants recruited from a clinical setting, 59 (20.3%) were recruited from private physiotherapy clinics with the remaining three participants recruited from private pain management or psychology clinical settings.

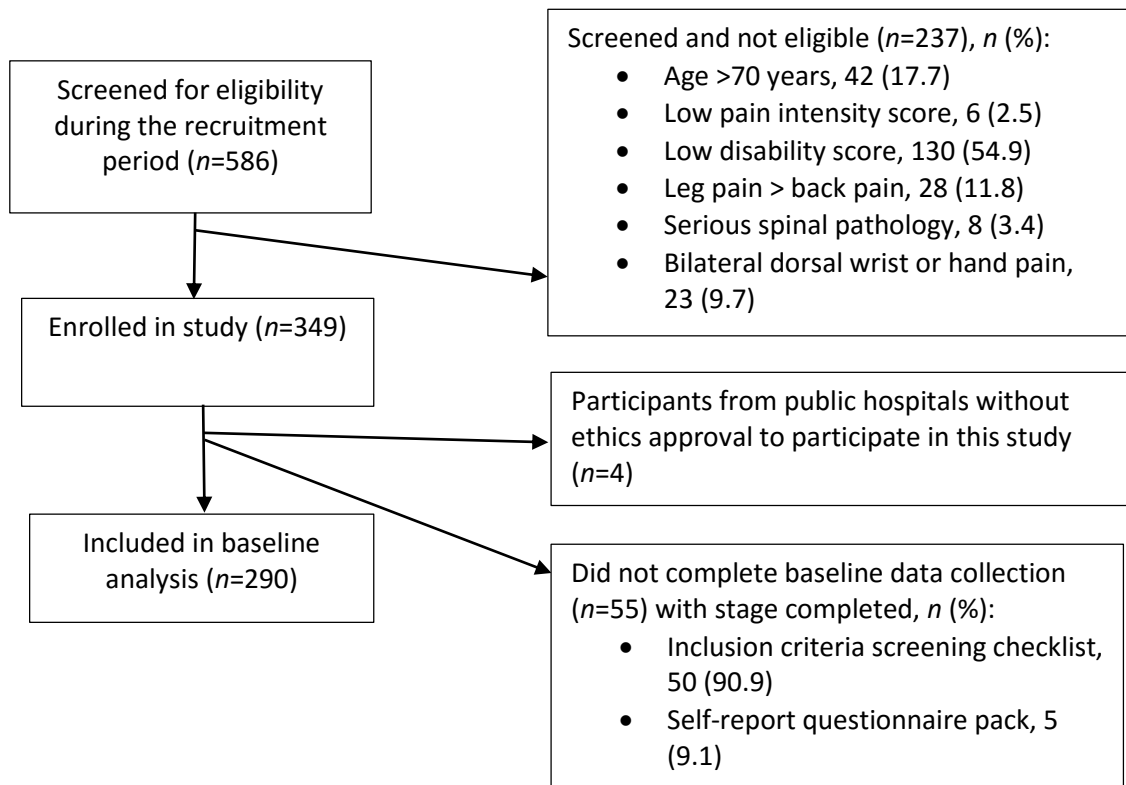


Figure 2.1. Flow of participants through the cross sectional study.

Table 2.1. Descriptive Characteristics of Potential Participants Screened but not Eligible

Demographic variable	Summary statistic (n=237)
Age (years), median (IQR)	51.0 (34.0-65.0)
Sex*, n (%) female	126 (55.5)
Reason for exclusion, n (%)	
Age > 70	42 (17.7)
Low pain intensity score (NRS)	6 (2.5)
Low disability score (RMDQ)	130 (54.9)
Leg pain > back pain	28 (11.8)
Serious spinal pathology	8 (3.4)
Bilateral dorsal wrist or hand pain	23 (9.7)

IQR = interquartile range; NRS = Numerical Rating Scale; RMDQ = Roland Morris Disability Questionnaire.

* 10 missing (n=227 for sex).

Table 2.2. Descriptive Characteristics of Participants who did and did not Complete Baseline Assessment

Demographic variable	Summary statistic		p-value*
	Non-completers (n=55)	Completers (n=290)	
Age (years), median (IQR)	48.0 (31.0-63.0)	50.5 (37.0-60.0)	0.326
Sex†, n (%) female	29 (61.7)	166 (57.2)	0.566
Pain intensity (NRS), mean (SD); range 0-10	6.1 (1.9)	5.8 (1.9)	0.282
Disability (RMDQ), median (IQR); range 0-24	9.0 (7.0-15.0)	9.0 (6.0-13.0)	0.149
Percent LBP (vs. leg pain), n (%)			0.667
100%	26 (47.3)	145 (50.0)	
80%	24 (43.6)	110 (38.0)	
60%	5 (9.1)	35 (12.1)	

IQR = interquartile range; NRS = Numerical Rating Scale; SD = standard deviation; RMDQ = Roland Morris Disability Questionnaire; LBP = low back pain.

* Independent t-test, Mann-Whitney U, or Chi Squared test depending on data type and distribution.

† 8 missing from the non-completers (n=47 for sex for the non-completers).

Table 2.3. Baseline Descriptive Characteristics of Included Participants – Total Cohort and Stratified by STarT Back Screening Tool Risk Subgroup*

Variable	Total cohort <i>n</i> =290†	Risk subgroup†			p-value
		Low <i>n</i> =82 (28.3%)	Medium <i>n</i> =116 (40.0%)	High <i>n</i> =92 (31.7%)	
Demographic variables					
Age (years)	50.5 (37.0-60.0)	49.0 (40.0-61.0)	51.5 (41.0-58.0)	50.0 (32.0-60.0)	0.556
Sex, <i>n</i> (%) female	166 (57.2)	49 (59.8)	73 (62.9)	44 (47.8)	0.079
Education level ¹ (years), mean (<i>SD</i>)	14.8 (3.6)	15.7 (3.9)	14.7 (3.3)	14.1 (3.5)	0.012^a
Employment status, <i>n</i> (%) yes working	223 (76.9)	65 (79.3)	89 (76.7)	69 (75.0)	0.799
Occupation ² , <i>n</i> (%)					≤0.001
Manual	72 (26.2)	13 (16.1)	24 (22.2)	35 (40.7)	
Sedentary	174 (63.3)	61 (75.3)	76 (70.4)	37 (43.0)	
Not working	29 (10.6)	7 (8.6)	8 (7.4)	14 (16.3)	
Compensation status ³ , <i>n</i> (%) yes compensated	45 (15.8)	15 (18.5)	15 (13.2)	15 (16.9)	0.712
Pain duration ⁴ (months)	120.0 (42.0-240.0)	120.0 (42.0-240.0)	120.0 (48.0-300.0)	120.0 (36.0-192.0)	0.481
Overall general health (COOP-WONCA)‡; range 1-5	3.0 (2.0-4.0)	2.0 (2.0-3.0)	3.0 (2.0-3.5)	3.0 (3.0-4.0)	≤0.001

Variable	Total cohort <i>n</i> =290†	Risk subgroup†			p-value
		Low <i>n</i> =82 (28.3%)	Medium <i>n</i> =116 (40.0%)	High <i>n</i> =92 (31.7%)	
Clinical measures					
Pain intensity (NRS), mean (<i>SD</i>); range 0-10	5.8 (1.9)	4.7 (1.8)	6.0 (1.8)	6.5 (1.6)	≤0.001
Disability (RMDQ); range 0-24	9.0 (6.0-13.0)	6.0 (5.0-8.0)	9.0 (7.0-13.0)	12.0 (8.5-15.0)	≤0.001
Percent LBP (vs. leg pain), <i>n</i> (%)					0.091
100%	145 (50.0)	51 (62.2)	49 (42.2)	45 (48.9)	
80%	110 (38.0)	23 (28.1)	50 (43.1)	37 (40.2)	
60%	35 (12.1)	8 (9.8)	17 (14.7)	10 (10.9)	

Variable	Total cohort <i>n</i> =290†	Risk subgroup†			p-value
		Low <i>n</i> =82 (28.3%)	Medium <i>n</i> =116 (40.0%)	High <i>n</i> =92 (31.7%)	
Psychological measures					
Depression (DASS-21); range 0-42	6.0 (2.0-14.0)	4.0 (2.0-6.0)	6.0 (2.0-12.0)	14.0 (6.0-24.0)	≤0.001
Anxiety (DASS-21); range 0-42	4.0 (2.0-8.0)	2.0 (0.0-6.0)	4.0 (2.0-8.0)	6.0 (2.0-11.0)	≤0.001 ^b
Stress (DASS-21); range 0-42	12.0 (6.0-20.0)	8.0 (4.0-14.0)	12.0 (6.0-18.0)	16.0 (10.0-24.0)	≤0.001
Fear avoidance beliefs-PA (FABQ-PA); range 0-24	15.0 (11.0-19.0)	13.0 (7.0-17.0)	13.5 (10.0-17.0)	18.0 (14.0-21.0)	≤0.001 ^c
Fear avoidance beliefs-W ⁵ (FABQ-W); range 0-42	16.0 (8.0-27.0)	13.0 (6.5-24.0)	15.0 (6.0-24.0)	20.0 (9.0-32.0)	0.010 ^c
Pain catastrophising ⁶ (PCS total); range 0-52	17.0 (9.0-27.0)	9.0 (4.0-15.0)	16.0 (10.0-23.0)	28.0 (20.5-36.0)	≤0.001
Rumination ⁶ (PCS); range 0-16	6.0 (3.0-10.0)	3.0 (1.0-7.0)	6.0 (3.0-9.0)	10.0 (5.0-13.5)	≤0.001
Magnification ⁶ (PCS); range 0-12	3.0 (1.0-5.0)	2.0 (1.0-4.0)	3.0 (1.0-5.0)	5.0 (4.0-7.0)	≤0.001
Helplessness ⁶ (PCS); range 0-24	8.0 (4.0-13.0)	4.0 (2.0-6.0)	7.0 (4.0-11.0)	13.0 (9.0-16.5)	≤0.001
Perceived risk of persistence (NRS 0-10)	9.0 (8.0-10.0)	8.0 (7.0-9.0)	9.0 (8.0-10.0)	9.0 (8.0-10.0)	≤0.001
Self-efficacy (PSEQ)§; range 0-60	42.0 (32.0-50.0)	49.0 (44.0-53.0)	43.0 (32.0-49.0)	34.0 (24.5-41.0)	≤0.001
Chronic pain acceptance (CPAQ-total)§; range 0-48	26.0 (21.0-31.0)	30.0 (26.0-36.0)	27.0 (23.0-31.0)	21.0 (15.5-25.5)	≤0.001
Chronic pain acceptance-pain willingness (CPAQ)§; range 0-24	9.0 (6.0-12.0)	11.0 (8.0-15.0)	9.0 (7.0-12.5)	6.0 (3.0-9.0)	≤0.001
Chronic pain acceptance-activity engagement (CPAQ)§; range 0-24	18.0 (14.0-21.0)	20.0 (17.0-22.0)	18.0 (15.0-21.0)	15.0 (11.0-19.0)	≤0.001

SD = standard deviation; NRS = Numerical Rating Scale; RMDQ = Roland Morris Disability Questionnaire; LBP = low back pain; DASS-21 = Depression Anxiety Stress Scale; FABQ-PA = Fear Avoidance Beliefs Questionnaire – Physical Activity; FABQ-W = Fear Avoidance Beliefs Questionnaire – Work; PCS = Pain Catastrophising Scale; PSEQ = Pain Self-efficacy Questionnaire; CPAQ = Chronic Pain Acceptance Questionnaire.

* Values represent median (interquartile range) unless otherwise indicated. Evaluation of subgroup differences were performed using a one-way Analysis of Variance (ANOVA), Kruskal-Wallis test, or Chi Squared test depending on data type and distribution. Where there are statistically significant differences between groups, group contrasts indicated a significant difference between all three subgroups unless otherwise indicated by ^{a b c}. Boldface indicates statistical significance ($p < 0.05$).

† $n = 290$ (total cohort), $n = 82$ (low risk group), $n = 116$ (medium risk group), and $n = 92$ (high risk group) except where there are missing values.

‡ COOP-WONCA charts lower scores indicate better health.

§ PSEQ higher scores indicate greater confidence.

§ CPAQ higher scores indicate greater acceptance.

Group contrasts:

Education level^a = significant difference between the low and the high risk group only (no significant difference between the low and the medium risk group; no significant difference between the medium and the high risk group).

DASS-21 anxiety score^b = significant differences between the low and the medium risk group and the low and the high risk group (no significant difference between the medium and the high risk group).

FABQ-PA/W scores^c = significant differences between the medium and the high risk group and the low and the high risk group (no significant difference between the low and the medium risk group).

Missing values:

Educational level¹ = 14 missing (6, 1, and 7 missing respectively from the low, medium, and high risk group).

Occupation² = 15 missing (1, 8, and 6 missing respectively from the low, medium, and high risk group).

Compensation status³ = 6 missing (1, 2, and 3 missing respectively from the low, medium, and high risk group).

Pain duration⁴ = 4 missing (0, 2, and 2 missing respectively from the low, medium, and high risk group).

Fear avoidance beliefs-work⁵ = 9 missing (2, 2, and 5 missing respectively from the low, medium, and high risk group).

Pain catastrophising⁶ = 1 missing (1 missing from the medium risk group).

The characteristics of the participants included in the analysis with respect to the total cohort and each SBST risk subgroup can be seen in Table 2.3. The SBST categorised 82 participants (28.3%) as low risk, 116 (40.0%) as medium risk, and 92 (31.7%) as high risk. The SBST risk subgroups did not significantly differ for the majority of the demographic variables. There was no significant difference ($p \geq 0.05$) in age, sex, employment status, compensation status, or pain duration between the three SBST risk subgroups. However, there was a significant difference ($p = 0.012$) in mean years of education. Group contrasts indicated that the high risk group had significantly fewer years in education compared to the low risk group. The SBST risk subgroups were significantly different ($p \leq 0.001$) with respect to occupation with a higher than expected proportion of participants with a manual occupation in the high risk group and a higher than expected proportion of participants with a sedentary occupation in the low risk group. There was a significant difference ($p \leq 0.001$) between the three SBST risk subgroups in the median score on the overall general health scale of the COOP-WONCA charts. There were increasingly higher scores (equating to poorer health) moving from the low to medium to high risk group.

There was a significant difference in pain intensity ($p \leq 0.001$) and self-reported disability ($p \leq 0.001$) between all three SBST risk subgroups with increasing pain and increasing disability in the low to medium to high risk group (Table 2.3). The SBST risk subgroups did not significantly differ ($p = 0.091$) with respect to the percentage of LBP versus leg pain.

With respect to the psychological measures, all three SBST risk subgroups were significantly different with greater negative psychological affect and cognitions, decreasing self-efficacy, and decreasing chronic pain acceptance in the low to medium to high risk group (Table 2.3). The exception to this was the DASS-21 anxiety score and the FABQ-PA/W scores where group contrasts indicated that only two of the three groups differed from one another. Group contrasts for the DASS-21 anxiety score indicated that the low risk group was significantly less anxious than the medium and the high risk group with no significant difference between the medium and the high risk group. On both the physical activity and work subscales of the FABQ, the high risk group scored significantly higher than both the medium and the low risk group, while there was no significant difference between the low and the medium risk group.

2.5 Key findings

The results of this study demonstrated that the risk profile of the SBST subgroups was related to clinical measures of pain and disability and to psychological factors. The SBST risk subgroups demonstrated increasingly higher levels of pain, disability, and negative psychological affect and cognitions in the low to medium to high risk group.

2.6 References

- ASGHARI, A. & NICHOLAS, M. 2001. Pain self-efficacy beliefs and pain behaviour. A prospective study. *Pain*, 94, 85-100.
- AUSTRALIAN BUREAU OF STATISTICS. 2013. *1220.0 - ANZSCO Australian and New Zealand standard classification of occupations, 2013, version 1.2* [Online]. Australian Bureau of Statistics 2013. Available: <http://www.abs.gov.au/ANZSCO> [Accessed 29 April 2014].
- BALAGUE, F., MANNION, A., PELLISE, F. & CEDRASCHI, C. 2012. Non-specific low back pain. *Lancet*, 379, 482-491.
- BENECIUK, J., ROBINSON, M. & GEORGE, S. 2015. Subgrouping for patients with low back pain: a multidimensional approach incorporating cluster analysis and the STarT Back Screening Tool. *The Journal of Pain*, 16, 19-30.
- CHOU, R., QASEEM, A., SNOW, V., CASEY, D., CROSS, T., SHEKELLE, P. & OWENS, D. 2007. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Annals of Internal Medicine*, 147, 478-491.
- COSTA, L., MAHER, C., MCAULEY, J., HANCOCK, M., HERBERT, R., REFSHAUGE, K. & HENSCHKE, N. 2009. Prognosis for patients with chronic low back pain: inception cohort study. *British Medical Journal*, 339, b3829.
- DIONNE, C., VON KORFF, M., KOESELL, T., DEYO, R., BARLOW, W. & CHECKOWAY, H. 2001. Formal education and back pain: a review. *Journal of Epidemiology and Community Health*, 55, 455-468.
- DUNN, K. & CROFT, P. 2005. Classification of low back pain in primary care: using "bothersomeness" to identify the most severe cases. *Spine*, 30, 1887-1892.
- DWORKIN, R., TURK, D., FARRAR, J., HAYTHORNTHWAIT, J., JENSEN, M., KATZ, N., KERNS, R., STUCKI, G., ALLEN, R., BELLAMY, N., CARR, D., CHANDLER, J., COWAN, P., DIONNE, R., GALER, B., HERTZ, S., JADAD, A., KRAMER, L., MANNING, D., MARTIN, S., MCCORMICK, C., MCDERMOTT, M., MCGRATH, P., QUESSY, S., RAPPAPORT, B., ROBBINS, W., ROBINSON, J., ROTHMAN, M., ROYAL, M., SIMON, L., STAUFFER, J., STEIN, W., TOLLETT, J., WERNICKE, J. & WITTER, J. 2005. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, 113, 9-19.
- FISH, R., MCGUIRE, B., HOGAN, M., MORRISON, T. & STEWART, I. 2010. Validation of the Chronic Pain Acceptance Questionnaire (CPAQ) in an internet sample and development and preliminary validation of the CPAQ-8. *Pain*, 149, 435-443.
- FOSTER, N., THOMAS, E., BISHOP, A., DUNN, K. & MAIN, C. 2010. Distinctiveness of psychological obstacles to recovery in low back pain patients in primary care. *Pain*, 148, 398-406.
- FRITZ, J., BENECIUK, J., GEORGE, S., HILL, J. & HAY, E. 2011. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Physical Therapy*, 91, 722-736.

- GATCHEL, R., PENG, Y., PETERS, M., FUCHS, P. & TURK, D. 2007. The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological Bulletin*, 133, 581-624.
- GEORGE, S., VALENCIA, C. & BENECIUK, J. 2010. A psychometric investigation of fear-avoidance model measures in patients with chronic low back pain. *Journal of Orthopaedic and Sports Physical Therapy*, 40, 197-205.
- HILL, J., DUNN, K., LEWIS, M., MULLIS, R., MAIN, C., FOSTER, N. & HAY, E. 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632-641.
- HILL, J., DUNN, K., MAIN, C. & HAY, E. 2010b. Subgrouping low back pain: a comparison of the STarT Back Tool with the Örebro Musculoskeletal Pain Screening Questionnaire. *European Journal of Pain*, 14, 83-89.
- JACKSON, T., WANG, Y., WANG, Y. & FAN, H. 2014. Self-efficacy and chronic pain outcomes: a meta-analytic review. *The Journal of Pain*, 15, 800-814.
- JACOB, T., BARAS, M., ZEEV, A. & EPSTEIN, L. 2001. Low back pain: reliability of a set of pain measurement tools. *Archives of Physical Medicine and Rehabilitation*, 82, 735-742.
- KAIVANTO, K., ESTLANDER, A., MONETA, G. & VANHARANTA, H. 1995. Isokinetic performance in low back pain patients: the predictive power of the Self-Efficacy Scale. *Journal of Occupational Rehabilitation*, 5, 87-99.
- KEELE UNIVERSITY. *STarT Back* [Online]. Available: <http://www.keele.ac.uk/sbst/startbacktool/> [Accessed 21 October 2015].
- KUIJER, W., BROUWER, S., DIJKSTRA, P., JORRITSMA, W., GROOTHOF, J. & GEERTZEN, J. 2005. Responsiveness of the Roland-Morris Disability Questionnaire: consequences of using different external criteria. *Clinical Rehabilitation*, 19, 488-495.
- LAISNE, F., LECOMTE, C. & CORBIERE, M. 2012. Biopsychosocial predictors of prognosis in musculoskeletal disorders: a systematic review of the literature (corrected and republished). *Disability and Rehabilitation*, 34, 1912-1941.
- LINTON, S. & BOERSMA, K. 2003. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. *The Clinical Journal of Pain*, 19, 80-86.
- LOVIBOND, S. & LOVIBOND, P. 1995. *Manual for the Depression Anxiety Stress Scales*, Sydney, Psychology Foundation.
- MANNICHE, C., ASMUSSEN, K., LAURITSEN, B., VINTERBERG, H., KREINER, S. & JORDAN, A. 1994. Low Back Pain Rating Scale: validation of a tool for assessment of low back pain. *Pain*, 57, 317-326.
- MCCRACKEN, L. & ECCLESTON, C. 2005. A prospective study of acceptance of pain and patient functioning with chronic pain. *Pain*, 118, 164-169.
- MCCRACKEN, L. & ECCLESTON, C. 2006. A comparison of the relative utility of coping and acceptance-based measures in a sample of chronic pain sufferers. *European Journal of Pain*, 10, 23-29.

- MCCRACKEN, L., VOWLES, K. & ECCLESTON, C. 2004. Acceptance of chronic pain: component analysis and a revised assessment method. *Pain*, 107, 159-166.
- MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.
- NICHOLAS, M. 2007. The Pain Self-efficacy Questionnaire: taking pain into account. *European Journal of Pain*, 11, 153-163.
- OSMAN, A., BARRIOS, F., GUTIERREZ, P., KOPPER, B., MERRIFIELD, T. & GRITTMANN, L. 2000. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. *Journal of Behavioral Medicine*, 23, 351-365.
- RABEY, M., SLATER, H., O'SULLIVAN, P., BEALES, D. & SMITH, A. 2015. Somatosensory nociceptive characteristics differentiate subgroups in people with chronic low back pain: a cluster analysis. *Pain*, 156, 1874-1884.
- ROLAND, M. & FAIRBANK, J. 2000. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine*, 25, 3115-3124.
- ROLAND, R. & MORRIS, R. 1983. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low back pain. *Spine*, 8, 141-144.
- SULLIVAN, M., BISHOP, S. & PIVIK, J. 1995. The Pain Catastrophizing Scale: development and validation. *Psychological Assessment*, 7, 524-532.
- SWINKELS-MEEWISSE, E., SWINKELS, R., VERBEEK, A., VLAEYEN, J. & OOSTENDORP, R. 2003. Psychometric properties of the Tampa Scale for Kinesiophobia and the Fear-avoidance Beliefs Questionnaire in acute low back pain. *Manual Therapy*, 8, 29-36.
- VAN TULDER, M., BECKER, A., BEKKERING, T., BREEN, A., GIL DEL REAL, M., HUTCHINSON, A., KOES, B., LAERUM, E. & MALMIVAARA, A. 2006. Chapter 3 European guidelines for the management of acute nonspecific low back pain in primary care. *European Spine Journal*, 15 (Supplement 2), S169-S191.
- VAN WEEL, C., KONIG-ZAHN, C., TOUW-OTTEN, F., VAN DUIJIN, N. & MEYBOOM-DE JONG, B. 2012. *Measuring functional status with the COOP/WONCA charts: a manual* [Online]. UMCG / University of Groningen, Research Institute. Available: http://www.rug.nl/research/share/research/tools/handleiding_coopwonca2edruk.pdf.
- WADDELL, G., NEWTON, M., HENDERSON, I., SOMERVILLE, D. & MAIN, C. 1993. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain*, 52, 157-168.
- WAI, E., HOWSE, K., POLLOCK, J., DORNAN, H., VEXLER, L. & DAGENAIS, S. 2009. The reliability of determining "leg dominant pain". *The Spine Journal*, 9, 447-453.
- WRIGHT, M., WREN, A., SOMERS, T., GOETZ, M., FRAS, A., HUH, B., ROGERS, L. & KEEFE, F. 2011. Pain acceptance, hope, and optimism: relationships to pain and adjustment in patients with chronic musculoskeletal pain. *The Journal of Pain*, 12, 1155-1162.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Chapter 3: Prospective Study

3.0 Aim

This study was a prospective design evaluating the predictive and discriminative ability of the STarT Back Screening Tool (SBST) for pain intensity, self-reported low back pain (LBP) related disability, and global self-perceived change at one year follow-up in an Australian population exclusively with chronic non-specific LBP.

It was hypothesised that: (i) the SBST risk subgroups would predict pain intensity, disability, and global perceived change at the one year follow-up with greater pain and disability and less self-perceived improvement as risk subgroup categorisation increased from low to medium to high; (ii) the SBST would discriminate between those participants who were and were not recovered with respect to pain, disability, and global perceived change; (iii) the SBST would have stronger predictive and discriminative ability for disability compared to pain and global perceived change; and (iv) the overall performance of the SBST would be attenuated in a population with chronic LBP in comparison to previous studies undertaken in populations with LBP of “mixed” or “variable” episode duration.

3.1 Methods

The baseline data collection procedures and variables were fully described in the Methods in Chapter 2 (section 2.2). Follow-up data were collected one year after entry into the study.

3.1.1 Participants

The participants were fully described in Chapter 2, section 2.2.1.

3.1.2 Procedures

At the end of one year, each participant was contacted via mail or email and asked to complete a follow-up questionnaire containing the follow-up measures (described below). The participants contacted by mail completed a paper copy of the follow-up questionnaire which was then returned to Curtin University, Perth, Western Australia in a reply paid envelope. The participants who did not respond to the first mailed follow-up questionnaire were sent a second questionnaire. If the participant failed to respond to the second mailed questionnaire, they were then contacted by telephone.

The majority of the participants were contacted via email and were provided with a link to an online version of the follow-up questionnaire (Qualtrics Research Suite software <https://curtin.asia.qualtrics.com/ControlPanel/>). Non-response to the first email resulted in a second email. The participants who did not respond to the second email were sent a paper copy of the questionnaire. Further non-response resulted in a second mailed paper copy of the follow-up questionnaire. If there was still no response, the participant was then contacted by telephone.

Data was exported from Qualtrics into an excel file and merged with the data recorded from the returned paper copies of the follow-up questionnaire. Participants were matched to their baseline data by their initials, date of birth, and postal code.

3.1.3 Follow-up measures

In line with recommended measures for research involving participants with chronic pain (Dworkin et al., 2005), the follow-up questionnaire included the following measures: (a) average pain intensity in the previous week measured on an 11-point Numerical Rating Scale (NRS) where zero indicates *no pain* and 10 the *worst pain imaginable* (Dworkin et al., 2005, Manniche et al., 1994), (b) LBP related disability measured by the Roland Morris Disability Questionnaire (RMDQ) (Kuijer et al., 2005, Roland and Fairbank, 2000, Roland and Morris, 1983), and (c) global perceived change measured by a single self-report item on a 7-point Global Rating of Change Scale (GRCS) (Dworkin et al., 2005). Adapted from Kamper and colleagues (2009), participants were asked: “With respect to your low back pain, how would you describe yourself now compared to one year ago when we examined you for the research project (laboratory session at Curtin University)?” The responses on the GRCS range from negative three (*very much worse*) to positive three (*very much improved*) (Dworkin et al., 2008). The SBST (<http://www.keele.ac.uk/sbst/startbacktool/>) (Keele University) was also readministered at the one year follow-up. Participants were also given the opportunity to outline any treatment they had in the previous year.

Each of the follow-up measures described above were dichotomised into “recovered” and “not recovered”. Not recovered with respect to pain was defined as a score of greater than or equal to three on the NRS (Mehling et al., 2015, Mehling et al., 2011). Not recovered with respect to disability was defined as a score of greater than or equal to seven on the RMDQ. This disability cut-off score was also used in the SBST’s original developmental and validation study (Hill et al., 2008) and in more recent studies that have investigated the

SBST (Hill et al., 2010b, Morsø et al., 2013b, Morsø et al., 2014, Mehling et al., 2015, Wideman et al., 2012). The use of the same threshold for disability as previous studies facilitated comparison with existing literature. Not recovered/not improved with respect to global perceived change was defined as a score of less than or equal to zero on the GRCS (i.e. *no change*, *minimally worse*, *much worse*, and *very much worse*). This cut-off score on the GRCS was selected to ensure participants who had not recovered were accurately captured. This threshold was thought to be appropriate for a sample with chronic LBP where any improvement, even if *minimally improved*, would be considered positive rather than negative.

3.1.4 Data analysis

The online survey software did not allow for missing values, prompting completion of all fields prior to allowing submission. If on the paper copy of the follow-up questionnaire the NRS for pain or the GRCS were incomplete, these measures were coded as missing. For paper copies of the SBST, if one item was missing the most common score (i.e. positive or negative) for the other items was used and if two or more items were missing, the SBST was coded as missing.

Comparisons between the participants who completed follow-up data collection with those who did not were made for demographic variables, baseline clinical measures (pain and disability), and baseline SBST risk subgroup stratification. Depending on data type and distribution, an independent sample *t*-test, Mann-Whitney *U*, or Chi Squared test was used to evaluate for differences between “non-completers” and “completers” for the aforementioned variables.

Descriptive statistics for the one year follow-up data were calculated for pain, disability, and global perceived change with respect to the total cohort and with respect to each baseline SBST risk subgroup. The SBST risk subgroup stratification at follow-up was also described and compared to the baseline stratification. The proportion of participants who were not recovered (as defined above) at the one year follow-up with respect to pain, disability, and global perceived change was calculated at a cohort level and by baseline SBST risk subgroup. Histograms and box plots were inspected for all continuous variables with normally distributed data expressed as means and standard deviations (*SD*). Skewed and nominal variables were expressed as medians and interquartile ranges (*IQR*). Frequency counts and percentages were used to express categorical variables.

In order to evaluate the predictive and discriminative ability of the SBST, a number of statistical methods were used. The predictive ability of the SBST was evaluated by calculating the additional risk (i.e. risk ratio [RR]) of non-recovery (as defined above) for participants classified by the SBST as medium risk or high risk at baseline. The low risk subgroup was used as the reference category. Comparisons between the medium and the high risk group were also made. A RR of less than 2.0 is unlikely to have much practical value, a RR of 3.0 can be considered a moderate effect, and a RR of 4.0 can be considered a strong effect (Ferguson, 2009).

In order to evaluate the accuracy of the SBST total score and the psychological subscale score to discriminate between participants who were recovered and not recovered at the one year follow-up with respect to pain, disability, and global perceived change, receiver operating characteristic (ROC) curves were constructed and the area under the curve (AUC) was calculated. The positive likelihood ratio (LR+), negative likelihood ratio (LR-), sensitivity, specificity, and the diagnostics odds ratio (DOR) for: (a) the low risk group versus the medium/high risk group and (b) the low/medium risk group versus the high risk group was calculated for the follow-up measures with significant AUC values. The SBST risk subgroups were collapsed into low/medium and medium/high as this reflects the risk subgroup cut-offs and facilitated comparison with previous studies (Hill et al., 2008, Mehling et al., 2015, Page et al., 2015). The DOR is a combined measure of discrimination and is calculated by dividing the LR+ by the LR- (Glas et al., 2003). As a single measure of test performance, the DOR was included to facilitate comparison of the results of this study with the results of previously published studies. The DOR ranges from zero to infinity (Glas et al., 2003). A higher DOR indicates better test discrimination, a value of one indicates that the test has no ability to discriminate, and a value lower than one indicates that the test incorrectly classifies (more negative test results amongst those with the condition or outcome of interest) (Glas et al., 2003). With respect to the AUC, an AUC of 0.50 suggests no discrimination, > 0.50 but < 0.70 poor discrimination, ≥ 0.70 but < 0.80 acceptable discrimination, ≥ 0.80 but < 0.90 excellent discrimination, and ≥ 0.90 outstanding discrimination (Hosmer et al., 2013). Based on the recommendations by Hosmer and colleagues (2013) and previously reported AUC values for disability and pain outcomes for other screening tools evaluated in populations with LBP, an AUC of less than 0.60 is considered non-informative (Traeger et al., 2015). Higher LR+ and lower LR- indicate better discrimination (Hill et al., 2008). Likelihood ratios above five or below 0.2 are generally

seen as supporting a strong test, whereas values close to one indicate poor test performance (Mehling et al., 2015).

All data analyses were performed using Stata 13.1 (Statacorp, Texas, USA) for Windows. Statistical significance was considered to be $p < 0.05$ for all analyses.

3.2 Ethics

This study was approved by the Curtin University Human Research Ethics Committee (Approval Number PT262/2013) (Appendix A). The longitudinal study from which this study obtained baseline data (Rabey et al., 2015) was also approved by the Curtin University Human Research Ethics Committee (Approval Number HR112/2012). All participants received a Participant Information Sheet in plain language outlining the purpose of the study, methods, risks and discomforts, and the benefits of the research (Appendix B). Confidentiality was guaranteed and participants were made aware that they were able to withdraw at any time without prejudice. Written informed consent was obtained from each participant (Appendix C).

3.3 Results

The flow of participants through the study is shown in Figure 3.1. One year follow-up data were available for 264 of the 290 (91.0%) participants enrolled in the study. Of the 26 (9.0%) participants lost to follow-up, three (11.5%) were from the low risk group, seven (26.9%) from the medium risk group, and 16 (61.5%) from the high risk group. There was no significant difference ($p > 0.05$) in age, sex, or baseline pain intensity between the participants who were lost to follow-up and the participants who completed the study (Table 3.1). However, the participants lost to follow-up had significantly higher baseline disability than the participants who completed the study (RMDQ median score of 10.5 versus 8.0, $p = 0.034$) (Table 3.1). The SBST subgroup proportions also differed significantly ($p = 0.002$) between the participants who did and did not complete the study with a higher proportion of non-completers (61.5%) than completers (28.8%) allocated to the high risk group (Table 3.1).

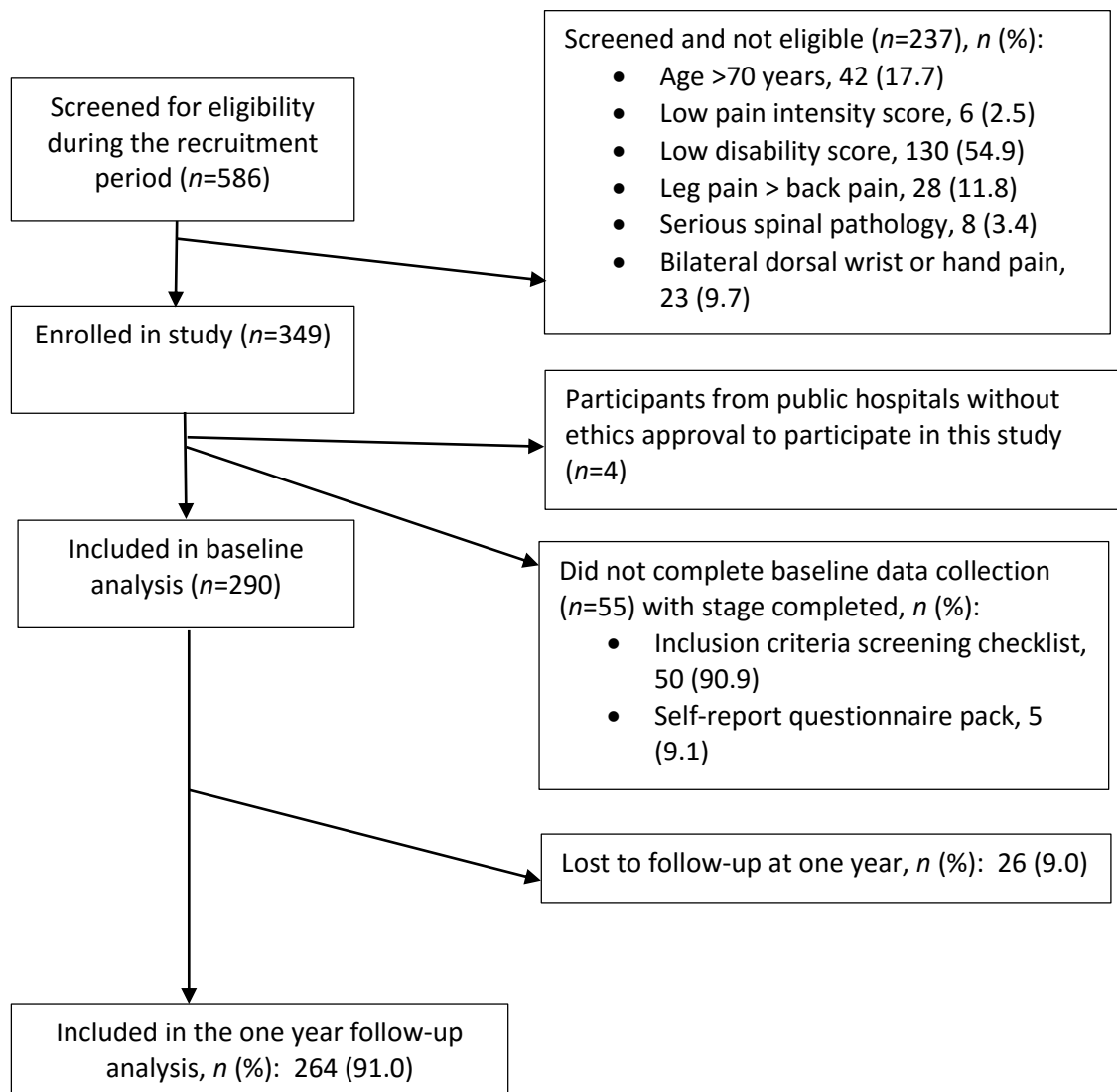


Figure 3.1. Flow of participants through the prospective study.

Table 3.1. Baseline Descriptive Characteristics of Participants who did and did not Complete Follow-up

Demographic variable	Summary statistic		p-value*
	Non-completers <i>n</i> =26 (9.0%)	Completers <i>n</i> =264 (91.0%)	
Age (years), median (IQR)	49.0 (33.0-59.0)	51.0 (39.0-60.0)	0.416
Sex, <i>n</i> (%) female	11 (42.3)	155 (58.7)	0.107
Pain intensity (NRS), median (<i>SD</i>); range 0-10	5.8 (1.7)	5.8 (1.9)	0.938
Disability (RMDQ), median (IQR); range 0-24	10.5 (7.0-15.0)	8.0 (6.0-12.5)	0.034
Baseline SBST risk subgroup, <i>n</i> (%)			0.002
Low	3 (11.5)	79 (29.9)	
Medium	7 (26.9)	109 (41.3)	
High	16 (61.5)	76 (28.8)	

IQR = interquartile range; NRS = Numerical Rating Scale; *SD* = standard deviation; RMDQ = Roland Morris Disability Questionnaire; SBST = STarT Back Screening Tool. Boldface indicates statistical significance ($p < 0.05$).

* Independent *t*-test, Mann-Whitney *U*, or Chi Squared test depending on data type and distribution.

Table 3.2. Baseline vs. Follow-up STarT Back Screening Tool Risk Subgroup Stratification, *n* (row %)

Baseline risk subgroup	Follow-up risk subgroup			Total
	Low	Medium	High	
Low	65 (82.3)	7 (8.9)	7 (8.9)	79 (100.0)
Medium	72 (66.1)	23 (21.1)	14 (12.8)	109 (100.0)
High	31 (40.8)	29 (38.2)	16 (21.1)	76 (100.0)
Total	168 (63.6)	59 (22.4)	37 (14.0)	264 (100.0)

When comparing SBST stratification at baseline to that at the one year follow-up, there were more participants in the low risk group and less in the medium risk and the high risk group at follow-up. Specifically, at follow-up there were 168 (63.6%) participants in the low risk group, 59 (22.4%) in the medium risk group, and 37 (14.0%) in the high risk group (Table 3.2). Comparatively, at baseline there were 82 (28.3%), 116 (40.0%), and 92 (31.7%) participants respectively in the low, medium, and high risk group (Table 2.3). Looking closer at the stability of the SBST risk subgroups over one year, 82.3% ($n=65$) of the low risk group at baseline remained in the low risk group at the one year follow-up with 8.9% ($n=7$)

moving into the medium risk group and 8.9% ($n=7$) moving into the high risk group. In the medium risk group, 78.9% ($n=86$) changed risk category at follow-up with 66.1% ($n=72$) moving into the low risk group and 12.8% ($n=14$) moving into the high risk group. In the high risk group, 78.9% ($n=60$) changed risk category at follow-up with 38.2% ($n=29$) moving into the medium risk group and 40.8% ($n=31$) moving into the low risk group (Table 3.2).

Table 3.3. Descriptive Characteristics at One Year Follow-up – Total Cohort and Stratified by Baseline STarT Back Screening Tool Risk Subgroup

Follow-up measure	Total cohort $n=264$	Risk subgroup*		
		Low $n=79$	Medium $n=109$	High $n=76$
Pain (NRS), mean (SD); range 0-10	4.2 (2.1)	3.5 (2.0)	4.4 (2.1)	4.5 (2.2)
n (%) not recovered (≥ 3)	201 (76.1)	51 (64.6)	88 (80.7)	62 (81.6)
Disability (RMDQ), median (IQR); range 0-24	4.0 (2.0-8.0)	3.0 (1.0-5.0)	5.0 (2.0-8.0)	6.0 (3.0-10.5)
n (%) not recovered (≥ 7)	83 (31.4)	12 (15.2)	38 (34.9)	33 (43.4)
Global perceived change[†] (GRCS), median (IQR); range -3 to 3 [‡]	1.0 (0.0-2.0)	1.0 (0.0-2.0)	1.0 (0.0-2.0)	0.5 (-1.0-2.0)
n (%) not improved (≤ 0)	117 (44.5)	31 (39.7)	48 (44.0)	38 (50.0)

NRS = Numerical Rating Scale; SD = standard deviation; RMDQ = Roland Morris Disability Questionnaire; IQR = interquartile range; GRCS = Global Rating of Change Scale.

* Lost to follow-up by risk subgroup = 3 from the low risk group, 7 from the medium risk group, and 16 from the high risk group.

† 1 missing from the low risk group ($n=263$ and $n=78$ for the low risk group).

‡ Where -3 equates to *very much worse* and 3 equates to *very much improved*.

At a total cohort level, 76.1% ($n=201$) of the participants were not recovered with respect to pain (NRS greater than or equal to three), 31.4% ($n=83$) were not recovered with respect to disability (RMDQ greater than or equal to seven), and 44.5% ($n=117$) were not improved as measured by the GRCS (less than or equal to zero) at the one year follow-up (Table 3.3). When stratified by baseline SBST subgroup, 64.6% ($n=51$) of the low risk group, 80.7% ($n=88$) of the medium risk group, and 81.6% ($n=62$) of the high risk group were not recovered with respect to pain. With respect to disability, 15.2% ($n=12$) of the low risk group, 34.9% ($n=38$) of the medium risk group, and 43.4% ($n=33$) of the high risk group

were considered disabled at follow-up. With respect to global perceived change, 39.7% ($n=31$) of the low risk group, 44.0% ($n=48$) of the medium risk group, and 50.0% ($n=38$) of the high risk group rated themselves as unchanged or worse (Table 3.3). A total of 215 (81.4%) participants reported receiving some form of intervention or treatment during the study period.

The RRs for pain, disability, and global perceived change at the one year follow-up for the medium and the high risk SBST subgroup, using the low risk group as the reference category, are shown in Figure 3.2. With respect to pain, the RR [95% confidence interval (CI)] was 1.25 [1.04, 1.51], $p=0.013$ for the medium risk group and 1.26 [1.03, 1.52], $p=0.020$ for the high risk group. The participants in both the medium and the high risk group had a 25% increased risk of being not recovered with respect to pain compared to the low risk group. When using the medium risk group as the reference category, the proportion of participants in the medium and the high risk group not recovered with respect to pain at the one year follow-up was similar with a RR of 1.01 [0.88, 1.16], $p=0.885$ (not shown in Figure 3.2).

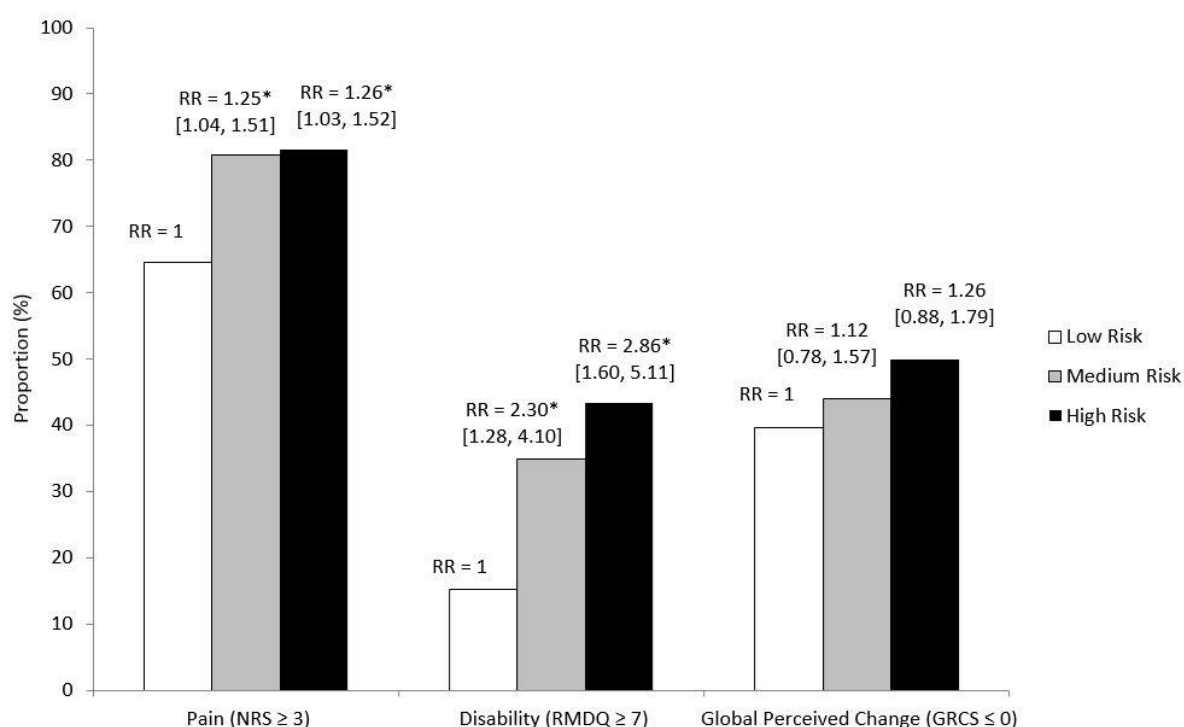


Figure 3.2. Relative risk of non-recovery at the one year follow-up by SBST risk subgroup. Risk ratios [95% confidence interval]. The low risk subgroup is the reference category. SBST = STarT Back Screening Tool; RR = risk ratio; NRS = Numerical Rating Scale; RMDQ = Roland Morris Disability Questionnaire; GRCS = Global Rating of Change Scale. * $p<0.05$.

Participants in the medium risk group (RR 2.30 [1.28, 4.10], $p=0.003$) and the high risk group (RR 2.86 [1.60, 5.11], $p\leq 0.001$) at baseline had a significantly increased risk of being considered disabled at the one year follow-up compared to the low risk group. The participants in the medium risk group had a 130% increased risk and the participants in the high risk group had a 190% increased risk of being considered disabled compared to the low risk group. The high risk group did not have a significantly higher risk of being considered disabled compared to the medium risk group with a RR of 1.25 [0.87, 1.79], $p=0.239$ (not shown in Figure 3.2).

Although a higher proportion of both the medium and the high risk group perceived themselves as not improved at the one year follow-up compared to the low risk group, the difference in risk was not significant with a RR for the medium risk group of 1.12 [0.78, 1.57], $p=0.558$ and a RR for the high risk group of 1.26 [0.88, 1.79], $p=0.201$. The high risk group was also not at an increased risk of not having improved at the one year follow-up compared to the medium risk group with a RR of 1.14 [0.83, 1.55], $p=0.424$ (not shown in Figure 3.2).

The ability of the SBST's total score and psychological subscale score at baseline to discriminate between participants who had and had not recovered at the one year follow-up with respect to each of the follow-up measures is shown in Table 3.4. The respective ROC curves are shown in Figure 3.3. The SBST's discriminative ability was the highest for disability with an AUC [95% CI] of 0.71 [0.64, 0.77] for the total score and 0.67 [0.60, 0.73] for the psychological subscale score. The AUC for pain was lower at 0.63 [0.55, 0.71] for both the total and the psychological subscale score. The ability of the SBST to discriminate global perceived change was the lowest with an AUC of 0.56 [0.49, 0.63] for the total score and 0.55 [0.48, 0.62] for the psychological subscale score.

Table 3.4. Discriminative Ability of the STarT Back Screening Tool Baseline Total Score and Psychological Subscale Score using the AUC

Follow-up measure	Case definition	n (%)	STarT Back Screening Tool AUC [95% CI]	
			Total score	Psychological subscale score
Pain	NRS ≥ 3	201 (76.1)	0.63 [0.55, 0.71]	0.63 [0.55, 0.71]
Disability	RMDQ ≥ 7	83 (31.4)	0.71 [0.64, 0.77]	0.67 [0.60, 0.73]
Global perceived change*†	GRCS ≤ 0	117 (44.5)	0.56 [0.49, 0.63]	0.55 [0.48, 0.62]

AUC = area under the curve; 95% CI = 95% confidence interval; NRS = Numerical Rating Scale; RMDQ = Roland Morris Disability Questionnaire; GRCS = Global Rating of Change Scale.

* 1 missing from the low risk group ($n=117/263$).

† Where -3 equates to *very much worse* and 3 equates to *very much improved*.

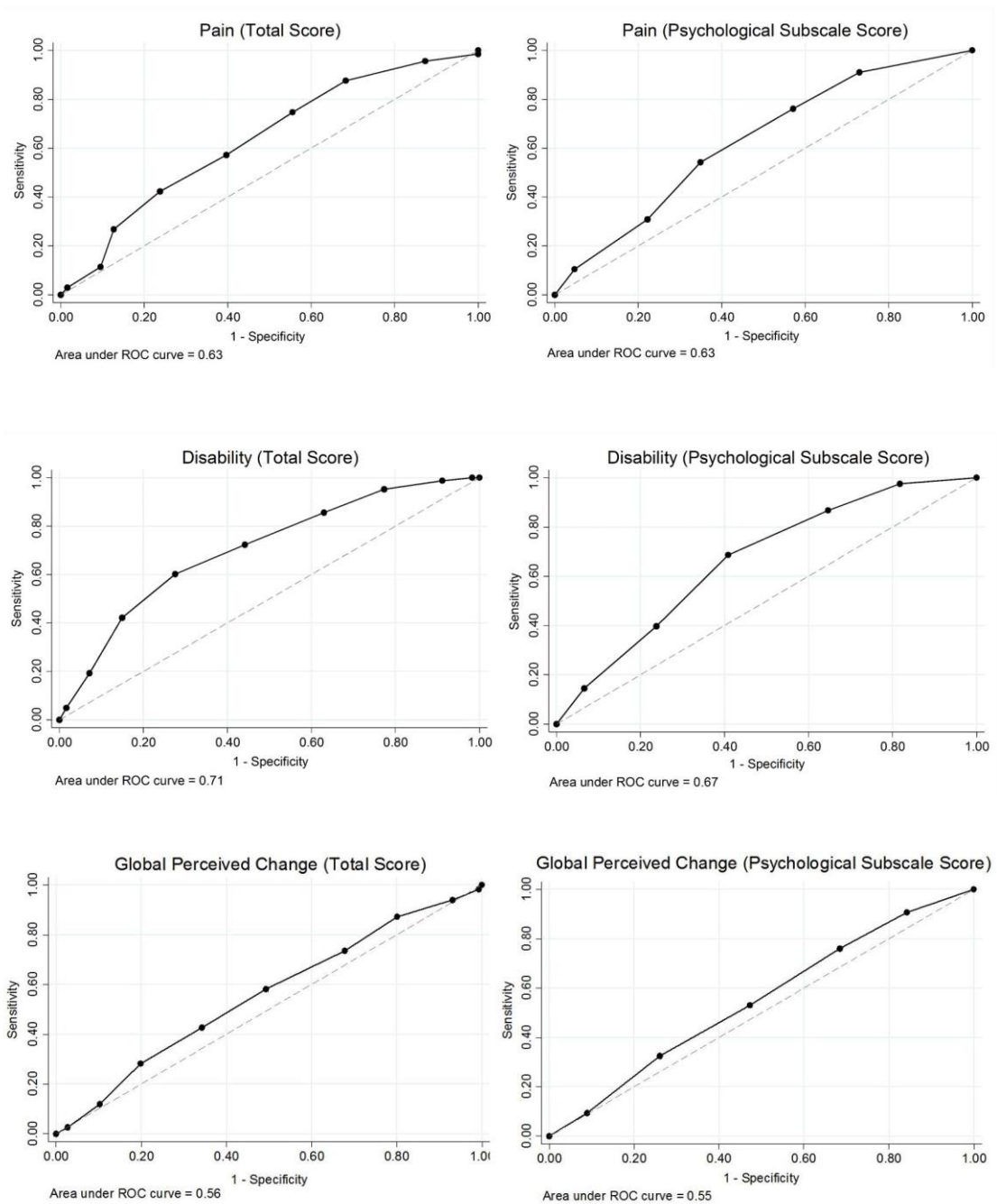


Figure 3.3. ROC curves for the SBST baseline total score and psychological subscale score for pain (NRS ≥ 3), disability (RMDQ ≥ 7), and global perceived change (GRCS ≤ 0) at the one year follow-up. ROC = receiver operating characteristic; SBST = STarT Back Screening Tool; NRS = Numerical Rating Scale; RMDQ = Roland Morris Disability Questionnaire; GRCS = Global Rating of Change Scale.

A score of less than or equal to one on the GRCS (i.e. *minimally improved, no change, minimally worse, much worse, and very much worse*) has been recommended to indicate a not recovered/not improved status (Dworkin et al., 2005, Dworkin et al., 2008). However, this study selected a cut-off score of less than or equal to zero on the GRCS to indicate non-recovery. This choice of cut-off score was rationalised in section 3.1.3. If this study had selected a cut-off score of less than or equal to one on the GRCS, the proportion of participants not recovered/not improved at the one year follow-up would have been higher with 66.2% ($n=174$) of the total cohort and 61.5% ($n=48$), 65.1% ($n=71$), and 72.4% ($n=55$) of the low, medium, and high risk group respectively not recovered/not improved. The RRs if selecting a cut-off score of less than or equal to one on the GRCS and using the low risk group as the reference category would have been: RR medium risk group 1.06 [0.85, 1.32], $p=0.614$ and RR high risk group 1.18 [0.94, 1.47], $p=0.153$. Using the medium risk group as the reference category: RR high risk group 1.11 [0.91, 1.35], $p=0.299$. The discriminative ability (AUC) of the SBST for global perceived change using less than or equal to one on the GRCS: 0.45 [0.38, 0.52] for the SBST total score and 0.43 [0.36, 0.51] for the psychological subscale score. Hence, the results of this study would be similar regardless of which cut-off score on the GRCS was selected.

The likelihood ratios, sensitivity, specificity, and DORs for the SBST risk subgroups for pain and disability at the one year follow-up are shown in Table 3.5. With respect to pain, the LR+ [95% CI] was 1.34 [1.06, 1.70] and the LR- [95% CI] was 0.57 [0.40, 0.82] for the low risk group versus the medium/high risk group with a sensitivity of 74.6%, specificity of 44.4%, and DOR of 2.35. The LR+ was 1.39 [0.84, 2.30] and the LR- was 0.89 [0.76, 1.05] for the low/medium risk group versus the high risk group with a sensitivity of 30.9%, specificity of 77.8%, and DOR of 1.56. With respect to disability, the LR+ was 1.36 [1.18, 1.57] and the LR- was 0.39 [0.22, 0.68] for the low risk group versus the medium/high risk group with a sensitivity of 85.5%, specificity of 37.0%, and DOR of 3.48. The LR+ was 1.67 [1.15, 2.43] and the LR- was 0.79 [0.65, 0.96] for the low/medium risk group versus the high risk group with a sensitivity of 39.8%, specificity of 76.2%, and DOR of 2.12.

Table 3.5. Discriminative Ability of the STarT Back Screening Tool Risk Subgroups - Likelihood Ratios, Sensitivity, Specificity, and the Diagnostic Odds Ratio

Subgroups	LR+ [95% CI]	LR- [95% CI]	Sensitivity (%)	Specificity (%)	DOR
Pain					
(NRS ≥ 3)					
L vs. M/H	1.34 [1.06, 1.70]	0.57 [0.40, 0.82]	74.6	44.4	2.35
L/M vs. H	1.39 [0.84, 2.30]	0.89 [0.76, 1.05]	30.9	77.8	1.56
Disability					
(RMDQ ≥ 7)					
L vs. M/H	1.36 [1.18, 1.57]	0.39 [0.22, 0.68]	85.5	37.0	3.48
L/M vs. H	1.67 [1.15, 2.43]	0.79 [0.65, 0.96]	39.8	76.2	2.12

LR+ = positive likelihood ratio; 95% CI = 95% confidence interval; LR- = negative likelihood ratio; DOR = diagnostic odds ratio; NRS = Numerical Rating Scale; L = low risk group; M = medium risk group; H = high risk group; RMDQ = Roland Morris Disability Questionnaire.

3.4 Key findings

The results of this study demonstrated that the SBST had moderate predictive and acceptable discriminative ability for disability at the one year follow-up. The SBST's predictive and discriminative ability was relatively weak for future pain and the tool was unable to identify those participants who perceived themselves as improved versus not improved at the one year follow-up.

3.5 References

- DWORKIN, R., TURK, D., FARRAR, J., HAYTHORNTHTWAITE, J., JENSEN, M., KATZ, N., KERNS, R., STUCKI, G., ALLEN, R., BELLAMY, N., CARR, D., CHANDLER, J., COWAN, P., DIONNE, R., GALER, B., HERTZ, S., JADAD, A., KRAMER, L., MANNING, D., MARTIN, S., MCCORMICK, C., MCDERMOTT, M., MCGRATH, P., QUESSY, S., RAPPAPORT, B., ROBBINS, W., ROBINSON, J., ROTHMAN, M., ROYAL, M., SIMON, L., STAUFFER, J., STEIN, W., TOLLETT, J., WERNICKE, J. & WITTER, J. 2005. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, 113, 9-19.
- DWORKIN, R., TURK, D., WYRWICH, K., BEATON, D., CLEELAND, C., FARRAR, J., HAYTHORNTHTWAITE, J., JENSEN, M., KERNS, R., ADER, D., BRANDENBURG, N., BURKE, L., CELLA, D., CHANDLER, J., COWAN, P., DIMITROVA, R., DIONNE, R., HERTZ, S., JADAD, A., KATZ, N., KEHLET, H., KRAMER, L., MANNING, D., MCCORMICK, C., MCDERMOTT, M., MCQUAY, H., PATEL, S., PORTER, L., QUESSY, S., RAPPAPORT, B., RAUSCHKOLB, C., REVICKI, D., ROTHMAN, M., SCHMADER, K., STACEY, B., STAUFFER, J., VON STEIN, T., WHITE, R., WITTER, J. & ZAVISIC, S. 2008. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *The Journal of Pain*, 9, 105-121.
- FERGUSON, C. 2009. An effect size primer: a guide for clinicians and researchers. *Professional Psychology - Research and Practice*, 40, 532-538.
- GLAS, A., LIJMER, J., PRINS, M., BONSEL, G. & BOSSUYT, P. 2003. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*, 56, 1129-1135.
- HILL, J., DUNN, K., LEWIS, M., MULLIS, R., MAIN, C., FOSTER, N. & HAY, E. 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632-641.
- HILL, J., DUNN, K., MAIN, C. & HAY, E. 2010b. Subgrouping low back pain: a comparison of the STarT Back Tool with the Örebro Musculoskeletal Pain Screening Questionnaire. *European Journal of Pain*, 14, 83-89.
- HOSMER, D., LEMESHOW, S. & STURDIVANT, R. 2013. *Wiley series in probability and statistics: applied logistic regression (3rd edition)*, New York, NY, USA, John Wiley & Sons.
- KAMPER, S., MAHER, C. & MACKAY, G. 2009. Global Rating of Change Scales: a review of strengths and weaknesses and considerations for design. *Journal of Manual and Manipulative Therapy*, 17, 163-170.
- KEELE UNIVERSITY. *STarT Back* [Online]. Available: <http://www.keele.ac.uk/sbst/startbacktool/> [Accessed 21 October 2015].
- KUIJER, W., BROUWER, S., DIJKSTRA, P., JORRITSMA, W., GROOTHOF, J. & GEERTZEN, J. 2005. Responsiveness of the Roland-Morris Disability Questionnaire: consequences of using different external criteria. *Clinical Rehabilitation*, 19, 488-495.
- MANNICHE, C., ASMUSSEN, K., LAURITSEN, B., VINTERBERG, H., KREINER, S. & JORDAN, A. 1994. Low Back Pain Rating Scale: validation of a tool for assessment of low back pain. *Pain*, 57, 317-326.

- MEHLING, W., AVINS, A., ACREE, M., CAREY, T. & HECHT, F. 2015. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19, 439-446.
- MEHLING, W., GOPISETTY, V., ACREE, M., PRESSMAN, A., CAREY, T., GOLDBERG, H., HECHT, F. & AVINS, A. 2011. Acute low back pain and primary care: how to define recovery and chronification? *Spine*, 36, 2316-2323.
- MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.
- MORSØ, L., KENT, P., MANNICHE, C. & ALBERT, H. 2014. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine Journal*, 23, 120-128.
- RABEY, M., SLATER, H., O'SULLIVAN, P., BEALES, D. & SMITH, A. 2015. Somatosensory nociceptive characteristics differentiate subgroups in people with chronic low back pain: a cluster analysis. *Pain*, 156, 1874-1884.
- ROLAND, M. & FAIRBANK, J. 2000. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine*, 25, 3115-3124.
- ROLAND, R. & MORRIS, R. 1983. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low back pain. *Spine*, 8, 141-144.
- TRAEGER, A., HENSCHKE, N., HUBSCHER, M., WILLIAMS, C., KAMPER, S., MAHER, C., MOSELEY, G. & MCAULEY, J. 2015. Development and validation of a screening tool to predict the risk of chronic low back pain in patients presenting with acute low back pain: a study protocol. *BMJ Open*, 5, e007916.
- WIDEMAN, T., HILL, J., MAIN, C., LEWIS, M., SULLIVAN, M. & HAY, E. 2012. Comparing the responsiveness of a brief, multidimensional risk screening tool for back pain to its unidimensional reference standards: the whole is greater than the sum of its parts. *Pain*, 153, 2182-2191.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Chapter 4: Discussion

4.1 Summary of key findings

This was the first large study to investigate the utility of the STarT Back Screening Tool (SBST) in a population exclusively with chronic non-specific low back pain (LBP) and the first to evaluate the tool in an Australian context. This study aimed to: (i) profile the SBST risk subgroups with respect to demographic variables, pain intensity, disability, and psychological measures and (ii) evaluate the SBST's predictive and discriminative ability for pain intensity, self-reported LBP related disability, and global self-perceived change at one year follow-up.

The results of the cross sectional study demonstrated that the risk profile of the SBST subgroups was related to clinical measures of pain and disability and to psychological factors. As hypothesised, the SBST risk subgroups demonstrated increasingly higher levels of pain, disability, and negative psychological affect and cognitions in the low to medium to high risk group.

The results of the prospective study demonstrated that the SBST had moderate predictive and acceptable discriminative ability for disability at the one year follow-up. As hypothesised, the SBST's predictive and discriminative ability was weaker for future pain. The tool was unable to identify those participants who perceived themselves as improved versus not improved at the one year follow-up.

This study adds to the body of research around the utility of the SBST. The results provide valuable information to clinicians on the usefulness and limitations of using the SBST in patients with chronic LBP in an Australian setting. The SBST has value as a substitute for the administration of multiple full-length, unidimensional questionnaires for first-line screening to help identify patients presenting with higher levels of pain, disability, and negative psychological affect and cognitions. The SBST provides an acceptable indication of future disability.

The results of this study (i.e. the "current study") in comparison to previously published studies will be discussed with comparisons focused on, but not limited to, the SBST's original developmental and validation study in a UK primary care setting (Hill et al., 2008) and the two studies undertaken in populations with predominantly (Morsø et al., 2014) or

exclusively (Page et al., 2015) chronic LBP. The strengths, limitations, and clinical implications of the current study along with recommendations for future research will also be discussed.

4.2 Baseline sample characteristics

In order to interpret and compare the results of the current study with previously published studies, it is necessary to examine more closely the recruitment method and the baseline characteristics of the sample including the SBST risk subgroup proportions and the demographic and clinical profile of the risk subgroups.

4.2.1 Recruitment considerations and baseline risk subgroup proportions

Participants in the current study were self-selecting volunteers primarily from the general community ($n=228$ [78.6%]) with 59 of the remaining 62 (20.3%) participants recruited from private physiotherapy clinics. Other studies that have evaluated the SBST have recruited care seeking patients with LBP from specific clinical settings. These clinical settings have included General Practice (GP) (Hill et al., 2008, Mehling et al., 2015, Morsø et al., 2013b, Morsø et al., 2013a, Morsø et al., 2016, Von Korff et al., 2014, Wideman et al., 2012), physiotherapy (Beneciuck et al., 2014, Beneciuk et al., 2013, Beneciuk et al., 2015, Fritz et al., 2011, George and Beneciuk, 2015, Morsø et al., 2013b, Morsø et al., 2016), chiropractic (Field and Newell, 2012, Kongsted et al., 2015, Kongsted et al., 2011, Morsø et al., 2016, Newell et al., 2015), or secondary care (Morsø et al., 2013a, Morsø et al., 2014, Morsø et al., 2016). Only one of the studies reviewed recruited more generally from the community through newspaper advertisements (Page et al., 2015). In the current study, 81.4% ($n=215$) of the participants reported having received some form of intervention or treatment during the study period. This indicates that the majority of the participants in the current study were care seekers however, they were not recruited from one specific care setting.

To be included in the current study, participants were required to have dominant axial LBP and were only included if they had 60% or more LBP versus leg pain. This criterion was used to reduce the likelihood of including participants with radiculopathy or radicular leg pain and therefore increase the likelihood of recruiting a sample with non-specific LBP. The original development and validation of the SBST was in a group who had non-specific LBP (Hill et al., 2008). The majority of studies that have evaluated the SBST have included participants with leg pain but did not appear to have used specific criteria related to the

relative proportion of LBP versus leg pain (Beneciuck et al., 2014, Beneciuk et al., 2013, Beneciuk et al., 2015, Field and Newell, 2012, Fritz et al., 2011, George and Beneciuk, 2015, Hill et al., 2008, Kongsted et al., 2015, Kongsted et al., 2011, Morsø et al., 2013b, Morsø et al., 2014, Newell et al., 2015). Hence, this sample may differ from other studies with respect to the nature of any referred leg pain.

A large number of the volunteers screened for inclusion were not eligible to participate, predominantly due to a low disability score (less than or equal to five) on the Roland Morris Disability Questionnaire (RMDQ), $n=130$ (54.9%) (Table 2.1). In the literature reviewed, only Wideman and colleagues (2012) limited inclusion to participants with clinically meaningful levels of pain (Numerical Rating Scale [NRS] score greater than three) and disability (RMDQ score greater than or equal to seven) making the current study somewhat unique in its inclusion criteria. In the current study, a large number of “low risk” participants may have been excluded relative to other studies due to the inclusion criteria used. A comparison of the baseline SBST risk subgroup proportions in the current study and in previously published studies follows and is supplemented by a summary of this data which can be found in Appendix D.

In studies that have investigated the SBST in a population with LBP of a “mixed” or “variable” episode duration (i.e. participants reported acute, subacute, or chronic LBP), the baseline subgroup proportions have ranged from 26%-59% for the low risk group, 29%-48% for the medium risk group, and 11%-28% for the high risk group (Beneciuk et al., 2015, Beneciuk et al., 2013, Field and Newell, 2012, Fritz et al., 2011, George and Beneciuk, 2015, Hill et al., 2008, Hill et al., 2011, Kongsted et al., 2011, Morsø et al., 2011, Morsø et al., 2013b, Newell et al., 2015). In studies that have investigated the SBST in a population with predominantly or exclusively chronic LBP, the baseline subgroup proportions have ranged from 28%-66% for the low risk group, 21%-33% for the medium risk group, and 13%-40% for the high risk group (Morsø et al., 2014, Page et al., 2015). In a study that investigated an item set analogous to the SBST in a population exclusively with acute LBP, the subgroup proportions were 32%, 46%, and 22% respectively for the low, medium, and high risk group (Mehling et al., 2015).

The current study had a relatively lower proportion of participants in the low risk group (28%) and a relatively higher proportion in the medium risk group (40%) and the high risk group (32%) at baseline compared to Hill and colleagues (2008) (developmental and

external sample: 40% and 47% respectively in the low risk group; 34% and 37% respectively in the medium risk group; 25% and 15% respectively in the high risk group). However, the SBST risk subgroup proportions in the current study were comparable to the subgroup proportions reported in the randomised controlled trial (RCT) (26%, 46%, and 28% respectively in the low, medium, and high risk group) undertaken in a UK primary care setting which evaluated stratified care using the SBST (Hill et al., 2011). This RCT recruited participants in the same manner and from the same population as Hill and colleagues (2008), yet the SBST risk subgroup proportions were different in these two studies. The SBST risk subgroup proportions in the current study were also comparable to Morsø and colleagues (2014) (28%, 33%, and 40% respectively in the low, medium, and high risk group), a study performed in Denmark in a secondary care setting where approximately 80% of the participants reported chronic LBP. Page and colleagues (2015) recruited participants exclusively with chronic LBP in a similar manner to the current study (i.e. community newspaper advertisements) but reported a larger proportion in the low risk group (66%) and a smaller proportion in the medium risk group (21%) and the high risk group (13%) in comparison to the current study. In summary, baseline SBST risk subgroup proportions varied amongst studies. This is not surprising and suggests differences in cohorts across countries and settings, regardless of the duration of LBP reported by the participants. In the current study, the SBST risk subgroups were well represented and the subgroup proportions did not demonstrate extremes in comparison to other studies.

Table 4.1. Baseline Descriptive Characteristics across Key Studies

Variable	Current study <i>n</i> =290	Hill et al., 2008 Developmental sample <i>n</i> =131	Hill et al., 2008 External sample <i>n</i> =500	Hill et al., 2011 RCT sample <i>n</i> =851	Morsø et al., 2014 <i>n</i> =960	Page et al., 2015 <i>n</i> =53
Age (years)	50.5 (37.0-60.0) (median [IQR])	44.0 (10.0) (mean [SD])	45.0 (9.7) (mean [SD])	49.8 (14.8)* (mean [SD])	52.0 (14.1) (mean [SD])	44.1 (13.3) (mean [SD])
Sex, <i>n</i> (%) female	166 (57.2)	77 (58.8)	293 (58.6)	500 (58.8)	521 (54.3)	23 (43.4)
Working, <i>n</i> (%) yes	246 (89.5) [†]	95 (72.5)	370 (74.0)	524 (61.6)	–	38 (71.7)
Pain duration (months)	120.0 (42-240) [†] (median [IQR]) Range 3-720	–	–	–	–	130.7 (112.0) (mean [SD]) Range 4-360
Pain duration stratified (months), <i>n</i> (%)	3-6: 8 (2.8) 7-36: 63 (22.0) >36: 215 (75.2) [†]	4-6: 15 (11.5) 7-36: 29 (22.1) >36: 34 (26.0)	4-6: 77 (15.4) 7-36: 125 (25.0) >36: 112 (22.4)	4-6: 123 (14.5) 7-36: 209 (24.6) >36: 178 (20.9)	≈ 80% chronic (> 3 months)	–
Pain (NRS); range 0-10	5.8 (1.9) (mean [SD])	–	–	5.3 (2.2)* (mean [SD])	6.0 (4.0-7.0) (median [IQR])	2.8 (2.3) (mean [SD])
Disability (RMDQ); range 0-24	9.0 (6.0-13.0) (median [IQR])	8.6 (6.6) (mean [SD])	9.1 (5.9) (mean [SD])	9.8 (5.7)* (mean [SD])	60.9 (44.0-78.0) [‡] (median [IQR])	18.4% (10.3%) [§] (mean [SD])
Leg pain, <i>n</i> (%)	145 (50.0)	75 (57.3)	303 (60.6)	530 (62.3)	–	–

Variable	Current study <i>n</i> =290	Hill et al., 2008 Developmental sample <i>n</i> =131	Hill et al., 2008 External sample <i>n</i> =500	Hill et al., 2011 RCT sample <i>n</i> =851	Morsø et al., 2014 <i>n</i> =960	Page et al., 2015 <i>n</i> =53
SBST total score; range 0-9	5.0 (3.0-7.0) (median [IQR])	4.3 (2.6) (mean [<i>SD</i>])	3.8 (2.3) (mean [<i>SD</i>])	–	–	–
SBST psych subscale score; range 0-5	3.0 (1.0-4.0) (median [IQR])	2.2 (1.6) (mean [<i>SD</i>])	1.8 (1.5) (mean [<i>SD</i>])	–	–	–
Low risk, <i>n</i> (%)	82 (28.3)	52 (39.7)	234 (46.8)	221 (26.0)	252 (27.7)	35 (66.0)
Medium risk, <i>n</i> (%)	116 (40.0)	45 (34.4)	186 (37.2)	394 (46.3)	296 (32.5)	11 (20.8)
High risk, <i>n</i> (%)	92 (31.7)	33 (25.2)	74 (14.8)	236 (27.7)	363 (39.9)	7 (13.2)

IQR = interquartile range; *SD* = standard deviation; – = data not available; NRS = Numerical Rating Scale; RMDQ = Roland Morris Disability Questionnaire; SBST = STarT Back Screening Tool.

* Weighted average and *SD*.

† Working = 15 missing; pain duration = 4 missing.

‡ 100-point RMDQ used which equates to 14.6 on the 24-point RMDQ.

§ Oswestry Disability Questionnaire used.

4.2.2 Baseline demographic and clinical characteristics

Baseline descriptive characteristics for a number of key studies that have evaluated the SBST in comparison to the current study are shown in Table 4.1. The key studies included in Table 4.1 are the SBST's original developmental and validation study (Hill et al., 2008), the SBST RCT (Hill et al., 2011), and the two studies undertaken in populations with predominantly (Morsø et al., 2014) or exclusively (Page et al., 2015) chronic LBP. The age of the participants, the proportion of females, and the proportion of working participants in the current study were similar to the studies described in Table 4.1.

In the current study, the SBST risk subgroups did not show statistically significant differences for the majority of the demographic variables including age, sex, and pain duration (Table 2.3). Similar results were reported by Beneciuk and colleagues (2015), Field and Newell (2012), and Kongsted and colleagues (2011). However, Morsø and colleagues (2014) reported a significant difference between the SBST risk subgroups for sex and pain duration but not for age. In the current study, the SBST risk subgroups also did not show statistically significant differences for employment status (Table 2.3). Similar results were reported by Beneciuk and colleagues (2015).

The current study was the first to evaluate for differences between the SBST risk subgroups for compensation status. There was no significant difference between the SBST risk subgroups on compensation status with only 15.8% ($n=45$) of the total cohort having a positive compensation status (Table 2.3). Conflicting results are reported in the literature with respect to the relationship between compensation status and outcomes in LBP disorders. A cross sectional comparison of compensation involvement versus no compensation involvement demonstrated an adverse effect on self-reported pain, depression, and disability for the compensation group (Rainville et al., 1997). A large prospective study in a primary care setting in Australia reported that the presence of compensation was independently associated with a longer recovery time (Henschke et al., 2008). Similarly, in a systematic review, Hayden and colleagues (2009) reported a consistent association between the presence of compensation and poorer outcomes. In contrast to the aforementioned, a systematic review by Ramond and colleagues (2011) reported that compensation issues did not independently influence outcomes.

In the current study, there were significant differences in education level, occupation, and self-reported overall general health between the SBST risk subgroups (Table 2.3). To this

investigator's knowledge, SBST risk subgroup differences on these demographic variables have not been previously reported. The high risk group had fewer years in education than the low risk group. A higher proportion of the high risk group had a manual occupation and a higher proportion of the low risk group had a sedentary occupation. Self-reported overall general health was increasingly poor across the three risk groups; the high risk group reported the poorest general health. There are likely to be many potential factors, interactions, and mechanisms to explain these findings that are well beyond the scope of this research. However, these findings are broadly in line with other studies. Costa and colleagues (2009) reported that patients with chronic LBP with a lower level of education were more likely to have a delayed recovery. A less favourable course of LBP amongst persons with less formal education has also been reported in an early review (Dionne et al., 2001). Dionne and colleagues (2001) postulated that level of education may have an influence on pain and disability because lower socio-economic status (education being a surrogate measures of socio-economic status) may be associated with factors such as physically demanding occupations and differences in access to or use of healthcare. Education level may also influence health literacy. Access to health services and health literacy may have implications for implementing targeted treatment pathways but speculating further on this is beyond the scope of this research. Other studies have reported that higher physical work demands were prognostic of an unfavourable outcome in LBP (Hayden et al., 2009, Shaw et al., 2001) and were a barrier to return to work (Shaw et al., 2011). In a systematic review of LBP prognosis, poorer general health was reported to be associated with poorer outcomes (Hayden et al., 2009).

In the current study, the baseline mean pain intensity of the total cohort was similar to that reported by Hill and colleagues (2011) and Morsø and colleagues (2014) (Table 4.1). The baseline median disability score for the total cohort was also comparable to that reported by Hill and colleagues (2008 and 2011) (Table 4.1). In contrast, in comparison to the current study, the secondary care cohort with predominantly chronic LBP showed higher disability at baseline (Morsø et al., 2014) (Table 4.1). Despite the inclusion criteria for the current study stipulating a minimum score of five on the RMDQ, the sample had a relatively low level of disability compared to normative data reported for a sample with chronic LBP recruited from a multidisciplinary tertiary-referral pain management center in Australia (Nicholas et al., 2008). The study by Page and colleagues (2015) reported relatively low baseline pain and disability (Table 4.1) however, the small sample size ($n=53$) in this study may have influenced these results. The proportion of the sample in the current study that

reported leg pain was similar to that reported in studies by Hill and colleagues (2008 and 2011) (Table 4.1).

4.3 Cross sectional profile of the STarT Back Screening Tool risk subgroups

The aim of the cross sectional study was to describe the distribution of clinical and psychological measures in relation to the SBST risk subgroups. Specifically, this study aimed to examine whether the three SBST risk subgroups differed with respect to pain, disability, and psychological profile. The results of this study demonstrated that from a statistical perspective, the SBST risk subgroups were successful in distinguishing elevated levels of pain, disability, and negative psychological affect and cognitions with reference to unidimensional questionnaires. To gain an appreciation of the clinical utility of the SBST, consideration should be given to whether or not these subgroup differences can be considered clinically relevant or meaningful in addition to being statistically different. In order to make this judgement, a number of factors related to the median (or mean) scores on the questionnaires for each of the SBST risk subgroups have been taken into consideration. These factors include qualitative descriptors, normative data in a clinical group with chronic LBP, the minimal clinically important difference (MCID) or minimal important change (MIC) where this information is available, and clinical opinion. As a preface to this discussion, this investigator acknowledges the limitations associated with MCID (or MIC) values and applying these values to baseline subgroups. The definition of a MCID varies in the literature with the common construct being a lowest amount of change that is important in some way (Beaton et al., 2002). The MCID typically relates to within-person change over time but could also relate to between-person differences (Beaton et al., 2002). However, other studies have looked at differences between people at one point in time (Beaton et al., 2002). There are a wide range of MCID values reported in the literature which vary based on the population studied and the methods used to derive these values (Wright et al., 2012, Beaton et al., 2002). These values are context specific rather than a fixed number (Beaton et al., 2002). Hence, caution is required when interpreting MCID values (Terwee et al., 2010). In the context of the following discussion, although MCID (or MIC) values are not typically used to compare groups on baseline values, they provide some perspective on subgroup differences observed in the current study.

4.3.1 Clinical measures of pain and disability

The results of the current study demonstrated increasingly higher pain and disability scores across the SBST risk subgroups with statistically significant differences between all three groups (Table 2.3). The mean pain intensity scores on the NRS for the SBST risk subgroups were 4.7, 6.0, and 6.5 respectively for the low, medium, and high risk group (Table 2.3). These scores suggest that all three risk subgroups had clinically meaningful levels of pain (Hush et al., 2009, Wideman et al., 2012). Normative pain data (as measured on a NRS) in a large Australian sample with chronic LBP has been reported as a mean (standard deviation [SD]) of 6.2 (2.0) (Nicholas et al., 2008). In the current study, the low risk group scored below this normative mean but the medium and the high risk group were close to this value. In LBP, the MIC on the NRS for pain has been reported as two (Ostelo et al., 2008). Given the aforementioned information, it is unlikely that the differences in pain score between the medium and the high risk group and the medium and the low risk group would be considered clinically meaningful. However, the difference in pain score between the low and the high risk group may be considered clinically meaningful.

The median scores on the RMDQ differed by three points between each of the three SBST risk subgroups with scores of 6.0, 9.0, and 12.0 respectively for the low, medium, and high risk group (Table 2.3). In comparison, the weighted mean RMDQ score was 4.5, 9.9, and 14.5 respectively for the low, medium, and high risk group in the SBST RCT (Hill et al., 2011). A median score of 14 on the RMDQ has been reported as normative in a large Australian sample with chronic LBP (Nicholas et al., 2008). Overall, the participants in the current study did not have a high level of disability at a cohort level or at an individual subgroup level. However, the median disability scores in the current study can be considered clinically meaningful particularly for the medium and the high risk group (Hush et al., 2009, Wideman et al., 2012). The MIC for the RMDQ has been reported as five points (Ostelo et al., 2008). Given the aforementioned information, the differences in disability between the low and the high risk group in the current study may be considered clinically meaningful.

The current study did not demonstrate a statistically significant difference between the three SBST risk subgroups for the percentage of LBP versus leg pain although the low risk group had the largest proportion of participants with 100% LBP (Table 2.3). Beneciuk and colleagues (2015) also reported no significant difference in symptom location (LBP versus leg pain) between the SBST risk subgroups. These findings are somewhat unexpected given that the presence of leg pain is a question in the SBST ("My back pain has spread down my

leg(s) at some time in the last two weeks”). Furthermore, in the original developmental and validation study, the SBST total score showed an excellent ability to discriminate the presence of referred leg pain (area under the curve [AUC] = 0.84) (Hill et al., 2008). In addition, a systematic review reported that the presence of radiating leg pain compared to LBP alone was associated with poorer outcomes (Konstantinou et al., 2013). The inclusion criteria for the current study may partially explain the results. In order to reduce the likelihood of participants with radiculopathy or radicular leg pain being included, participants were required to have dominant axial LBP (i.e. 60% or more LBP compared to leg pain). The result was 50% of the total cohort reported having 100% LBP; this profile was maintained across the risk subgroups.

4.3.2 Psychological measures

The results of the current study demonstrated increasingly higher scores across the SBST risk subgroups for the measures of depression, anxiety, stress, fear avoidance beliefs, pain catastrophising, and perceived risk of pain persistence (Table 2.3). The results also demonstrated lower scores for the measures of self-efficacy and chronic pain acceptance as SBST risk sub-group categorisation increased from low to medium to high (Table 2.3). There were statistically significant differences between all three risk subgroups for all psychological measures with the exception of the Depression Anxiety Stress Scale-21 (DASS-21) anxiety score and the Fear Avoidance Beliefs Questionnaire (FABQ) scores where the group contrasts indicated statistically significant differences between two groups only. Significant differences for the DASS-21 anxiety score were only found between the low risk group and the medium and the high risk group with no significant difference between the medium and the high risk group. The scores on the physical activity and work subscales of the FABQ demonstrated that the high risk group had significantly greater fear avoidance beliefs than both the medium and the low risk group with no significant difference between the low and the medium risk group. Although statistically significant, whether or not the subgroup differences on the psychological measures can be considered clinically relevant or meaningful is discussed below.

The current study was the first to use the DASS-21 as the reference questionnaire to measure depression, stress, and anxiety. The median depression scores on the DASS-21 were considered “normal” for both the low and the medium risk group and “moderate” for the high risk group (Lovibond and Lovibond, 1995) (Table 2.3). This suggests that the high risk group may have clinically meaningful higher levels of depression when compared to the

low and the medium risk group. The low and the medium risk group had median stress scores on the DASS-21 that were considered normal while the high risk group was considered “mildly” stressed (Lovibond and Lovibond, 1995) (Table 2.3). This suggests that there may be a small, but clinically relevant difference between the stress levels of the high risk group and the other two risk groups. With respect to the anxiety scores on the DASS-21, the total cohort (median score of 4.0 out of a maximum score of 42) and each of the SBST risk subgroups (median score of 2.0, 4.0, and 6.0 respectively for the low, medium, and high risk group) had low levels of anxiety with all scores considered normal (Lovibond and Lovibond, 1995) (Table 2.3). Given this information, a clinically meaningful difference in anxiety between the risk subgroups is unlikely. Reported normative data in a large Australian sample with chronic LBP were a median DASS score of 10.0 for depression, 6.0 for anxiety, and 13.0 for stress (Nicholas et al., 2008). In the current study, only the high risk group scored at or above these values.

In the current study, the median scores on the physical activity subscale of the FABQ were 13.0, 13.5, and 18.0 respectively for the low, medium, and high risk group (Table 2.3). The median scores on the work subscale of the FABQ were 13.0, 15.0, and 20.0 respectively for the low, medium, and high risk group (Table 2.3). It appears that the high risk group may have clinically meaningful greater fear avoidance beliefs than both the low and the medium risk group. This is consistent with where the statistically significant differences were found with the group contrasts (Table 2.3). In the current study, extracting information on fear avoidance beliefs using a single question in the SBST was not adequate to distinguish between the three risk subgroups. This is in keeping with the findings from Kent and colleagues (2014) who evaluated the concurrent validity of brief psychosocial screening questions against full-length reference questionnaires. These authors reported that for the construct of fear of movement, a combination of two questions rather than a single one resulted in a stronger correlation with its reference questionnaire (FABQ) (Spearman Correlation Coefficient = 0.926) (Kent et al., 2014). The current study did not compare individual SBST constructs/questions to their respective full-length reference questionnaires. The complexities and ambiguities around measuring the construct of fear has previously been highlighted (Lundberg et al., 2011).

In the current study, the median scores on the Pain Catastrophising Scale (PCS) were 9.0, 16.0, and 28.0 for the low, medium, and high risk group respectively (Table 2.3). According to the PCS User Manual, a total score of 20 is the 50th percentile (Sullivan, 2009). Only the

high risk group scored above the 50th percentile. Hill and colleagues (2008) used a score of greater than or equal to 20 on the PCS to indicate non-recovery. Taking this information into account, it would suggest that only the high risk group had relatively high levels of pain catastrophising. It appears that the high risk group may have clinically meaningful higher levels of pain catastrophising than both the low and the medium risk group.

The current study was the first to investigate the relationship between perceived risk of pain persistence and SBST stratification. The results demonstrated statistically significant differences between all three SBST risk subgroups on this variable (Table 2.3). However, the median score on the NRS for this measure was 8.0 for the low risk group and 9.0 for both the medium and the high risk group. This suggests that the differences between the three risk groups are unlikely to be clinically meaningful. The median duration of LBP for all three risk groups was 10 years. It is therefore not surprising that all groups scored very high on the NRS for the risk of pain persistence. Although this measure is not a construct found in the SBST, the question “I feel that my back pain is terrible and it’s never going to get any better” may in part reflect the construct of perceived risk of pain persistence.

The current study was also the first to investigate the relationship between the positive constructs of self-efficacy and chronic pain acceptance and the SBST risk subgroups. These constructs are not included in the SBST but the results of this study supported that the risk subgroups were inversely related to these positive constructs. There were statistically significant differences in self-efficacy and chronic pain acceptance (total score, pain willingness subscale, and activity engagement subscale) between the three risk subgroups. The high risk group had the lowest self-efficacy and chronic pain acceptance score. A relationship between lower self-efficacy and greater functional impairment, distress, and pain severity in chronic pain samples has been reported (Jackson et al., 2014). Lower chronic pain acceptance has been shown to be associated with greater distress and pain related disability in persistent chronic musculoskeletal pain disorders (McCracken and Eccleston, 2006, Wright et al., 2011). A normative median total score of 24.0 on the Pain Self Efficacy Questionnaire (PSEQ) has been reported in a large Australian sample with chronic LBP (Nicholas et al., 2008). In the current sample, all three SBST risk subgroups had a median score higher than 24.0 indicating that this sample had higher levels of self-efficacy than the sample in the study by Nicholas and colleagues (2008). Although the SBST risk subgroups were statistically different with respect to both of these positive constructs, it

appears that only the high risk group may have clinically meaningful lower levels of self-efficacy and chronic pain acceptance compared to the other two groups.

4.3.3 Summary - clinically meaningful differences between risk subgroups

Although statistically significant, the differences between the three SBST risk subgroups for the clinical measures of pain and disability and the psychological measures may not be clinically meaningful differences between all three subgroups. It appears that: (a) the high and the low risk subgroup may have clinically meaningful differences in pain and disability with the medium risk group approximating the high risk group; (b) the high risk group may have clinically meaningful higher levels of depression, stress, fear avoidance beliefs, and catastrophising compared to the other two groups; (c) all three risk subgroups had a similar low (normal) level of anxiety; (d) all three risk subgroups perceived a similar high risk of pain persistence; and (e) the high risk group may have clinically meaningful lower self-efficacy and chronic pain acceptance compared to the other two groups. Broadly, the high risk group appears to stand separate from the other two groups with respect to the psychological measures suggesting the tool can identify a low and a high contribution of psychological factors to a patient's presentation. The SBST may be more effective at distinguishing between two risk subgroups rather than three in a population with chronic LBP.

4.3.4 Comparisons to other studies

The results of the current cross sectional study are in keeping with the results from previously published cross sectional studies that have investigated the profile of the SBST risk subgroups. Studies that have investigated samples with LBP of variable duration in physiotherapy and chiropractic clinical settings have demonstrated that the SBST risk subgroups were related to pain intensity (Field and Newell, 2012, Fritz et al., 2011) and disability (Field and Newell, 2012, Fritz et al., 2011, Beneciuk et al., 2015). These studies reported increasing pain and disability from the low to medium to high risk group. However, these studies did not comment on whether or not the differences in pain and disability between the SBST risk subgroups were clinically meaningful in addition to being statistically different. In a cohort exclusively with chronic LBP, baseline SBST total score was correlated with baseline disability but not with baseline pain (Page et al., 2015). The absence of a correlation between the SBST total score and baseline pain intensity in this

study may have been due to the small sample size ($n=53$) or the low level of pain present at baseline (mean pain intensity 27.6%).

With respect to psychological measures, earlier studies that have investigated participants with LBP of variable duration have reported that depression (Beneciuk et al., 2015, Kongsted et al., 2011), fear avoidance beliefs (Beneciuk et al., 2015, Kongsted et al., 2011), catastrophising (Beneciuk et al., 2015, Kongsted et al., 2011), kinesiophobia (Beneciuk et al., 2015), and anxiety (Beneciuk et al., 2015) were related to the SBST risk subgroups. These studies also did not comment on whether or not the SBST risk subgroup differences were clinically meaningful in addition to being statistically different. Beneciuk and colleagues (2013) showed that the SBST total score and psychological subscale score were positively correlated with fear avoidance beliefs, catastrophising, kinesiophobia, and depression. Page and colleagues (2015) reported that the SBST total score was positively correlated with kinesiophobia.

The results of the current study suggest that the SBST may be more effective at distinguishing between two risk subgroups rather than three in a population with chronic LBP. Similar to the results of the current study, Beneciuk and colleagues (2015) reported the presence of two rather than three distinctive patterns among psychological and disability measures in a cohort with variable duration LBP in a physiotherapy setting in the USA.

In summary, the results from the current study are in keeping with the results from previously published studies. When evaluated in a cross sectional manner, pain, disability, and psychological measures have consistently demonstrated a relationship with the SBST risk subgroups and/or SBST score.

4.3.5 STarT Back Screening Tool as a substitute for full-length, unidimensional measures

The SBST was developed as a practical way to subgroup patients with LBP based on modifiable prognostic indicators which are appropriate targets for intervention. The practicality of the SBST is derived from its multidimensional nature and low responder and assessor burden making it feasible for use in busy clinical practice. The tool was designed to reduce the need to administer multiple full-length, unidimensional questionnaires in

order to determine risk profile for screening purposes and subsequently assist in care decisions. The SBST's acceptability as a surrogate measure for multiple full-length measures as first-line screening has been previously reported in samples with LBP of variable duration in a physiotherapy setting (Beneciuk et al., 2015) and a UK primary care setting (Wideman et al., 2012). In a cross sectional study, Beneciuk and colleagues (2015) reported that the SBST adequately distinguished between a number of full-length, unidimensional psychological and disability measures to identify elevated levels of psychological distress, maladaptive coping, and disability. These authors concluded that the SBST was a suitable replacement for the administration of multiple unidimensional questionnaires (Beneciuk et al., 2015). Wideman and colleagues (2012) demonstrated that the SBST was responsive to change at the four month follow-up and could be used instead of its reference standards to imply changes in pain severity, disability, fear of movement, and catastrophising therefore improving the efficiency of assessment and reassessment in a clinical setting. Although the study by Wideman and colleagues (2012) was prospective in nature, these results lend support to using the SBST for baseline screening assessment rather than multiple full-length, unidimensional questionnaires.

In keeping with previous studies, the results of the current cross sectional study support that the SBST has value as a substitute for the administration of multiple full-length, unidimensional questionnaires for first-line screening to assist in the identification of patients presenting with low or high levels of pain, disability, and negative psychological affect and cognitions. The clinical implications of this will be discussed in detail in section 4.8.

4.4 The one year follow-up sample characteristics

In order to interpret and compare the results of the current prospective study with previously published studies, it is necessary to examine more closely the follow-up sample characteristics. The low attrition rate at the one year follow-up ($n=26$ [9.0%]) is one of the strengths of this study. Although the participants that were lost to follow-up had statistically significant higher baseline disability (RMDQ median of 10.5 versus 8.0 for participants who completed the study) and were more likely to be from the high risk group (61.5% versus 28.8% for participants who completed the study), given the overall low attrition rate, this is unlikely to have influenced the results (Table 3.1).

4.4.1 STarT Back Screening Tool risk subgroup stratification at follow-up

At the one year follow-up there were 168 (63.6%) participants in the low risk group, 59 (22.4%) in the medium risk group, and 37 (14.0%) in the high risk group (Table 3.2). Comparatively, at baseline there were 82 (28.3%), 116 (40.0%), and 92 (31.7%) participants respectively in the low, medium, and high risk group (Table 2.3). Given that participants in the current study had a baseline LBP duration median of 10 years, it was unexpected to see such a large proportion in the low risk group and a relatively smaller proportion in the medium and the high risk group at the one year follow-up in comparison to baseline stratification. Although speculative, the reasons for this may have been related to the natural history of the disorder, treatment effects (81.4% [$n=215$] of the participants reported having received some form of treatment during the study period), or self-reporting inconsistencies. The fluctuating nature of LBP over time with respect to both pain and disability (Dunn et al., 2006) may have also influenced SBST stratification as the tool was only administered at baseline and at one follow-up time point.

In a cohort undergoing physiotherapy treatment, Beneciuk and colleagues (2013) reported a larger proportion of patients in the low risk group ($n=77$ [69.4%]) and a smaller proportion of patients in the medium risk ($n=23$ [20.7%]) and high risk group ($n=11$ [9.9%]) at the six month follow-up compared to baseline stratification (baseline stratification: low risk $n=53$ [36.3%], medium risk $n=55$ [37.7%], and high risk $n=38$ [26.0%]). These results are similar to those seen in the current study at the one year follow-up time point.

Other studies have reported on changes in SBST stratification at intake compared to a short time after receiving either chiropractic or physiotherapy treatment. Newell and colleagues (2015) showed that approximately one third of chiropractic patients changed SBST risk subgroups between the initial chiropractic visit and two days post initial visit with equal numbers improving or worsening. Similarly, Beneciuk and colleagues (2014) described SBST categorisation “change” following four weeks of physiotherapy treatment. In this study, the majority of patients “improved” ($n=60$ [48.8%]) or remained “stable” ($n=50$ [40.6%]) with only 13 (10.6%) participants “worse” as indicated by moving up a risk group or remaining in the high risk subgroup following four weeks of physiotherapy treatment (Beneciuk et al., 2014). The high risk group was the least stable as 81.8% of this group changed to low or medium risk from intake to four weeks post intake (Beneciuk et al., 2014). In the medium risk group, 76.0% changed with 71.7% moving to low risk (Beneciuk et al., 2014). In the low risk group, 11.3% changed, all to the medium risk group (Beneciuk

et al., 2014). These authors concluded that relying solely on intake SBST, particularly if initially categorised as high or medium risk, could be misleading as the majority of the participants improved after four weeks of physiotherapy (Beneciuck et al., 2014). The results of the study by Beneciuk and colleagues (2014) suggest that repeated screening may provide additional prognostic information and more accurately capture the risk profile of the patient. In the current study, short term stability of the SBST risk subgroups was not evaluated.

4.4.2 Non-recovery rates at follow-up

The results of the current study showed different rates of non-recovery for the total cohort for the follow-up measures of pain (NRS score greater than or equal to three; $n=201$ [76.1%]), disability (RMDQ score greater than or equal to seven; $n=83$ [31.4%]), and global perceived change (global rating of change scale [GRCS] less than or equal to zero; $n=117$ [44.5%]) (Table 3.3). This pattern of non-recovery was maintained when stratified by SBST risk subgroup. Whereas a large proportion of this cohort rated themselves as not recovered with respect to pain, a much smaller proportion rated themselves as not recovered with respect to disability. Neither the pain nor the disability domain seemed to easily translate into the participant's perception of improvement on a GRCS. The findings of the current study are comparable to George and Beneciuk (2015) who reported that non-recovery with respect to pain (NRS) was much higher than non-recovery with respect to disability (RMDQ) although the recovery criteria in this study were much more stringent than the current study (i.e. non-recovery defined as NRS score greater than 0 and RMDQ score of greater than or equal to three).

In the current study, the proportion of participants who rated themselves as improved on the GRCS was 50.0% ($n=38$) for the high risk group, 56.0% ($n=61$) for the medium risk group, and 60.3% ($n=47$) for the low risk group. In contrast, the perceived risk of pain persistence at baseline (on an 11-point NRS with 10 indicating a *very large risk*) was 9.0, 9.0, and 8.0 for the high, medium, and low risk group respectively. It appears that the participants in this study expected their pain to persist when evaluated at baseline yet a large proportion felt that they had improved one year later. This suggests that other factors informed perceived global change despite a high perceived risk of pain persistence at baseline and a high rate of non-recovery with respect to pain at the one year follow-up.

The construct of LBP recovery is complex, multidimensional, and highly individualised (Hush et al., 2009, Kamper et al., 2010). Measures of pain and disability may capture different aspects of recovery to that of the patient's perceived recovery (Kamper et al., 2010). It has been previously reported that pain and disability do not align with or easily translate into self-rated perceived recovery (Hush et al., 2009, Mehling et al., 2011). The results of the current study clearly support these statements.

The variability in the proportion of participants not recovered with respect to the three follow-up measures may also be related to the cut-off scores that were selected to indicate recovered versus not recovered. There is no consensus on a NRS for pain score or RMDQ score that equates to recovered or not recovered (Kamper et al., 2010). In fact, a systematic review that evaluated measures used to assess recovery from non-specific LBP found a striking lack of consistency amongst measures used, making comparisons across studies problematic (Kamper et al., 2011). Even when the domain measured was identical, it was often measured differently across studies (Kamper et al., 2011). The justification for selecting the cut-off scores for the current study is outlined in detail in section 3.1.3. In brief, the cut-off scores were considered clinically relevant and in line with those used in previous studies, therefore facilitating comparisons between this study and earlier publications on the SBST (Hill et al., 2008, Hill et al., 2010b, Mehling et al., 2015, Mehling et al., 2011, Morsø et al., 2013b, Morsø et al., 2014, Wideman et al., 2012). Inter-study comparisons are discussed below.

A number of studies have used greater than or equal to seven on the RMDQ (or equivalent) to indicate non-recovery at follow-up (Appendix E-H) (Hill et al., 2008, Mehling et al., 2015, Morsø et al., 2013b, Morsø et al., 2014, Page et al., 2015). A comparison of the current study with previously published studies on the proportion of participants considered disabled at follow-up, as indicated by greater than or equal to seven on the RMDQ, is summarised in Appendix E. In the current study, the proportion of participants considered disabled at follow-up was lower than that reported by Hill and colleagues (2008), Morsø and colleagues (2013b), and Morsø and colleagues (2014) at a total cohort level and at the level of each of the SBST risk subgroups (Appendix E). The proportion of participants considered disabled at follow-up in the high risk subgroup (43%) in the current study was particularly low compared to Hill and colleagues (2008) (78%), Morsø and colleagues (2013b) (64%), and Morsø and colleagues (2014) (80%) (Appendix E). In comparison to the current study, only the study by Mehling and colleagues (2015) that investigated a modified

SBST item set in a population exclusively with acute LBP, had a lower proportion of participants considered disabled at both the three month and the two year follow-up time points (Appendix E). In the only other study that has investigated the SBST in a population exclusively with chronic LBP, the proportion of participants not recovered was not reported for any of the follow-up measures at any of the follow-up time points (Page et al., 2015).

There are a number of factors that may account for the differences seen across studies with respect to the proportion of participants with future disability. These include baseline psychological profile, the timing of follow-up, the potential effects of treatment, cultural influences, and other factors related to the setting in which the study was undertaken. Baseline disability may have also influenced the level of disability at follow-up. As an example, in the secondary care cohort investigated by Morsø and colleagues (2014), baseline disability was higher than in the current study (Table 4.1). At follow-up, the proportion of participants considered disabled in the total cohort and in each SBST risk subgroup was also higher than in the current study (Appendix E).

4.5 Predictive and discriminative ability of the STarT Back Screening Tool

The SBST had better predictive and discriminative ability for future disability in comparison to future pain. The SBST was unable to identify those participants who had rated themselves as improved versus not improved at the one year follow-up. A comparison of the SBST's performance with respect to the follow-up measures of pain, disability, and global perceived change within the current study and in comparison to other studies will be discussed. Possible reasons for differences in the predictive and discriminative ability will be considered. The clinical implications of the findings from this study will be discussed in detail in section 4.8.

4.5.1 Pain and disability

The predictive ability of the SBST was evaluated by calculating the additional risk (i.e. risk ratio [RR]) of non-recovery for participants classified as medium or high risk at baseline with the low risk group used as the reference category. The results indicated that the high risk group (RR 2.86) and medium risk group (RR 2.30) had a meaningfully greater relative risk of being considered disabled at the one year follow-up compared to the low risk group (Figure 3.2). The high risk group (RR 1.26) and medium risk group (RR 1.25) had a statistically greater risk of non-recovery with respect to pain but this relative risk is unlikely to be of any

clinical significance. The lower RRs for pain may have been due to the relatively high proportion of participants ($n=51$ [64.6%]) who were not recovered in the low risk group; this would have attenuated the RRs because the low risk group was used as the reference category.

Two studies that have investigated the predictive ability of the SBST have reported RRs using the same RMDQ cut-off score as the current study to indicate future disability, as well as using the low risk group as the reference category (Morsø et al., 2013b, Morsø et al., 2014). A comparison of the predictive ability (using RRs) of the SBST for future disability between the current study, Morsø and colleagues (2013b), and Morsø and colleagues (2014) is summarised in Appendix F. The RRs for disability in the current study were higher than those reported by Morsø and colleagues (2014) (medium risk group RR 1.5 and high risk group RR 1.7) for the six month disability outcome in a secondary care setting in Denmark where participants reported predominantly chronic LBP. Morsø and colleagues (2014) compared the predictive ability of the SBST in the Danish secondary care setting to a Danish primary care setting with participants who reported variable duration LBP. The SBST performed better in the Danish primary care setting (medium risk group RR 2.3 and high risk group RR 3.5) when compared to the secondary care setting (Morsø et al., 2014). The weak predictive ability of the SBST demonstrated in the secondary care setting in Denmark may have been because the proportion of participants considered disabled at the six month follow-up was high at both a cohort and subgroup level with a particularly high proportion (almost 50%) of the low risk group considered disabled at follow-up (Morsø et al., 2014). Morsø and colleagues (2013b) investigated the predictive ability of the SBST for disability outcome at the three month follow-up in a Danish primary care setting. Participants were recruited from GP and physiotherapy clinics for this study and reported LBP of variable duration (Morsø et al., 2013b). The RR was 2.4 for the medium risk group and 2.7 for the high risk group (Morsø et al., 2013b). Morsø and colleagues (2013b) then compared the predictive ability of the SBST in the Danish primary care setting to that in a UK primary care setting. The RRs were higher in the UK primary care setting, especially for the high risk group (medium risk group RR 3.1 and high risk group RR 4.5) (Morsø et al., 2013b). In summary, the SBST showed better predictive ability for disability in the current study than in a Danish secondary care setting (Morsø et al., 2014), similar predictive ability to that reported in a Danish primary care setting (Morsø et al., 2013b, Morsø et al., 2014), and weaker predictive ability compared to a UK primary care setting as reported in the study by Morsø and colleagues (2013b).

In the current study, there was no significant difference in the risk of experiencing greater pain intensity or disability at the one year follow-up between the medium risk group and the high risk group (high risk group RR 1.01 for pain and 1.25 for disability). The distinction between medium and high risk stratification is made based on higher scores on the psychological subscale. Hence, it may be that the psychological subscale did not discriminate between the medium and the high risk subgroup in the current study. This may be linked to the cross sectional study results that demonstrated statistically significant but not necessarily clinically meaningful differences between the risk subgroups for the psychological measures. Alternatively, other potential confounding factors at baseline or positive treatment effects during the study period may help explain these results, however these were not explored in this study.

The results from the current study demonstrated that the SBST had better discriminative ability for disability than for pain at the one year follow-up (Table 3.4). The AUC for the SBST total score was 0.71 for disability, indicating acceptable discriminative ability. The AUC for the psychological subscale score was 0.67 for disability, indicating slightly poorer discrimination. Both the SBST total score and psychological subscale score had relatively poor discriminative ability for pain at the one year follow-up with an AUC of 0.63 for both scores. While other studies have evaluated the ability of the SBST total score to discriminate future measures of pain and disability, the current study is the only study to evaluate the discriminative ability of the psychological subscale score as well.

A comparison of the discriminative ability of the SBST total score (using the AUC) for future disability between the current study and previously published studies (Mehling et al., 2015, Morsø et al., 2013b, Morsø et al., 2014, Page et al., 2015) is summarised in Appendix G. The discriminative accuracy of the SBST total score for disability in the current study was similar to that reported in a Danish primary care setting (AUC 0.71) (Morsø et al., 2013b), a Danish primary care setting with participants recruited from physiotherapy clinics (AUC 0.73) (Morsø et al., 2014), and a Danish secondary care setting (AUC 0.69) (Morsø et al., 2014). However, the discriminative accuracy of the SBST total score was lower than that reported by Morsø and colleagues (2013b) in a UK primary care setting (AUC 0.81). In comparison to the current study, Page and colleagues (2015) reported that the SBST total score had an excellent ability to discriminate six month and 12 month disability outcome with an AUC [95% confidence interval (CI)] of 0.84 [0.69, 1.00] and 0.82 [0.61, 1.00]

respectively in a sample exclusively with chronic LBP. However, the confidence intervals were large which may have been due to the small sample size in the study. Page and colleagues (2015) measured disability with the Oswestry Disability Questionnaire (ODQ) using a cut-off score equivalent to greater than or equal to seven on the RMDQ which may also have influenced the results. In the study by Page and colleagues (2015), the SBST total score also had better discriminative ability for pain (case definition was greater than or equal to 37%) than the current study with an AUC of 0.73 [0.58, 0.86] for six month pain outcome and an AUC of 0.71 [0.54, 0.88] for 12 month pain outcome; the confidence intervals were also large.

Kongsted and colleagues (2015) investigated the discriminative ability of the SBST total score for pain and disability outcome in a sample predominantly with acute LBP; only 3% of the participants reported chronic LBP. The case definitions for poor outcome were stringent. A NRS for pain score equivalent to greater than zero and a RMDQ score equivalent to greater than two indicated a poor outcome. The SBST performed poorly as it was unable to discriminate pain cases at the three and the 12 month follow-up time points (Kongsted et al., 2015). The SBST also had a poor ability to discriminate future disability with an AUC of 0.59 [0.55, 0.63] and 0.60 [0.56, 0.64] respectively at the three month and the 12 month follow-up. Poor performance of the SBST in populations with acute LBP has been reported in other studies (Mehling et al., 2015, Morsø et al., 2016).

The discriminative ability of other screening tools for pain and disability outcomes in populations with LBP, using the AUC statistic, has been reported. The AUC in these studies has typically ranged from the low 0.60s to the high 0.80s (Gabel et al., 2011, Grotle et al., 2006, Hockings et al., 2008, Linton and Boersma, 2003, Nonclercq and Berquin, 2012). Based on the recommendations by Hosmer and colleagues (2013) and previously reported AUC values for disability and pain outcomes, an AUC of less than 0.60 can be considered non-informative (Traeger et al., 2015). Hence, the AUC values from the current study suggest that the SBST has value in discriminating future disability but is only somewhat informative in discriminating future pain in a population with chronic LBP. Consistent with other studies, the SBST demonstrated better discriminative ability for future disability.

In order to further evaluate the discriminative ability of the SBST for future pain and disability, the positive likelihood ratio (LR+), negative likelihood ratio (LR-), sensitivity, specificity, and the diagnostics odds ratio (DOR) for: (a) the low risk group versus the

medium/high risk group and (b) the low/medium risk group versus the high risk group was calculated (Table 3.5). With respect to pain, a positive test (membership in the medium/high risk group at baseline) resulted in a minimal increase (LR+ 1.34 [95% CI 1.06, 1.70]) and a negative test (not in the medium/high risk group at baseline) resulted in a minimal decrease (LR- 0.57 [95% CI 0.40, 0.82]) in the likelihood of being not recovered at the one year follow-up with a sensitivity of 74.6% (i.e. correctly identified as not recovered by a positive test result), specificity of 44.4% (i.e. correctly identified as recovered by a negative result), and DOR of 2.35. A positive test (membership in the high risk group at baseline) resulted in no significant increase (LR+ 1.39 [0.84, 2.30]) and a negative test (not in the high risk group at baseline) resulted in no significant decrease (LR- 0.89 [0.76, 1.05]) in the likelihood of being not recovered with respect to pain at the one year follow-up with a sensitivity of 30.9%, specificity of 77.8%, and DOR of 1.56.

With respect to disability, a positive test (membership in the medium/high risk group at baseline) resulted in a minimal increase (LR+ 1.36 [1.18, 1.57]) and a negative test (not in the medium/high risk group at baseline) resulted in a small decrease (LR- 0.39 [0.22, 0.68]) in the likelihood of being not recovered at the one year follow-up with a sensitivity of 85.5%, specificity of 37.0%, and DOR of 3.48. A positive test (membership in the high risk group at baseline) resulted in a minimal increase (LR+ 1.67 [1.15, 2.43]) and a negative test (not in the high risk group at baseline) resulted in a minimal decrease (LR- 0.79 [0.65, 0.96]) in the likelihood of being not recovered with respect to disability at the one year follow-up with a sensitivity of 39.8%, specificity of 76.2%, and DOR of 2.12.

In the current study, the LR+s were higher and the LR-s were lower for disability compared to pain, again indicating that the SBST performed better with respect to the disability measure. The overall performance of the SBST was the strongest for future disability for the low risk group versus the medium/high risk group with a DOR of 3.48 (Table 3.5). A higher DOR indicates better test discrimination (Glas et al., 2003). From a clinical practice perspective, a negative screening result (i.e. not being in the medium or the high risk subgroup) would have a moderate to high sensitivity (85.5%) to “rule out” an unfavourable disability outcome. High sensitivity is useful from an initial screening point of view where it is important to identify a potentially at risk patient for whom further assessment would be of value.

A comparison of the discriminative ability of the SBST (using LR+, LR-, sensitivity, specificity, and the DOR) for future disability between the current study and previously published studies (Hill et al., 2008, Mehling et al., 2015, Page et al., 2015) is summarised in Appendix H. In the original developmental and validation study by Hill and colleagues (2008), the likelihood ratios, sensitivity, specificity, and DORs for the SBST's subgroup cut-off scores to discriminate poor disability outcome at the six month follow-up were: (a) low risk group versus medium/high risk group LR+ 2.32 [1.96, 2.76], LR- 0.30 [0.23, 0.40], sensitivity 80.1%, specificity 65.4%, and DOR 7.73 and (b) low/medium risk group versus high risk group LR+ 5.51 [3.30, 9.28], LR- 0.74 [0.67, 0.81], sensitivity 29.6% (corrected value reported as noted by Mehling and colleagues (2015)), specificity 94.6%, and DOR 7.45. In comparison to Hill and colleagues (2008), the current study had likelihood ratios for disability that were lower for the LR+ and higher for the LR-. The DORs were correspondingly lower indicating that the SBST performed better in the UK primary care setting in a population with LBP of variable duration than in the current study. The likelihood ratios and DORs in the current study were marginally better than those reported by Mehling and colleagues (2015) in a study that evaluated a modified SBST item set in a sample where the participants reported acute LBP.

In summary, the SBST was better able to predict and discriminate future disability compared to future pain in the current study. This is in keeping with results reported in previously published studies. In comparison to a UK primary care setting where patients reported variable duration LBP, the SBST's performance was attenuated in the current study in a population with chronic LBP. Possible explanations for these results are considered in section 4.5.3. The clinical implications of these results are discussed in section 4.8.

4.5.2 Global perceived change

In the current study, the SBST was unable to identify participants who perceived themselves as improved versus not improved at the one year follow-up with the RRs not showing statistical significance (Figure 3.2) and a poor AUC with a CI that crossed 0.50 (Table 3.4). Other studies that have looked at the ability of the SBST to predict global perceived change or discriminate between those who perceived themselves as improved or not improved have reported similar results regardless of the setting, population, or follow-up period (Field and Newell, 2012, Newell et al., 2015, Page et al., 2015). Clearly the SBST

does not have utility as a predictor of this measure of recovery or non-recovery. Possible explanations for these results are discussed in section 4.5.3.

4.5.3 Reasons for the relatively weak performance of the STarT Back Screening Tool in the current study

In the current study, the SBST provided an acceptable indication of future disability but the tool was not as useful for identifying future pain and was unable to identify self-perceived global change. Previous publications have also demonstrated that the SBST's performance differed depending on the outcome measured. The SBST's performance has also differed across studies with the tool's predictive and discriminative ability overall weaker in this study compared to studies undertaken in a UK primary care setting.

There are a number of possible reasons as to why the SBST was relatively weak in its predictive and discriminative ability in the current study. These reasons include the follow-up measures evaluated, possible effects of treatment, the constructs included in the SBST, cross cultural influences on the tool's performance, the cut-off scores used to derive the SBST risk subgroups, the typical natural course of chronic LBP, the response format of the SBST, and consideration of the complex and multidimensional nature of chronic LBP.

The ability of a questionnaire to predict outcome is likely to be dependent on the outcome domain evaluated (Dworkin et al., 2008). The SBST was originally developed and validated to predict future disability so it is not surprising that the tool performed best with respect to this follow-up measure. Furthermore, the SBST includes items related to the pain and disability domains and therefore the tool can be expected to perform better with respect to these measures in comparison to global perceived change which is conceptually different from the instrument itself. This has also been demonstrated with the Örebro Musculoskeletal Pain Screening Questionnaire (Örebro). The discriminative ability of the Örebro for self-rated recovery has been shown to be poor in comparison to pain, disability, or work-related outcomes which are all constructs included in the Örebro (Heneweer et al., 2007, Jellema et al., 2007). A systematic review of the Örebro reported that the tool had moderate predictive ability for future pain, disability, and sick leave but poor predictive ability for self-reported global recovery (Hockings et al., 2008).

The participants in the current study were free to pursue any intervention or treatment they wanted. A large proportion of the sample (81.4% [$n=215$]) received some form of treatment during the study period. Given that the tool stratifies based on modifiable risk factors, if these factors were effectively treated, purposely or coincidentally, the predictive performance of the tool would decrease (Morsø et al., 2013b, Morsø et al., 2016).

The constructs included in the SBST were those shown to be the strongest independent predictors of future LBP related disability in a UK primary care setting where participants reported LBP of variable duration (Hill et al., 2008). The SBST's performance may have been weaker in the current study because the constructs included in the SBST may not be as important in chronic LBP. Although predictors of outcome are likely to be similar for acute and chronic LBP (Balague et al., 2012), it is likely that the specific factors involved, which combination of factors, or the strength of association of these factors with future outcomes are likely to differ depending on the duration of the LBP episode.

Cross cultural influences on the predictive performance of the SBST have been previously reported (Morsø et al., 2013b). The current study was the first to evaluate the SBST in an Australian context and this may have influenced the tool's performance in comparison to how the tool performed in a UK primary care setting where the SBST was initially developed and validated.

It may be that the cut-off scores used to derive the low, medium, and high risk subgroup are not as appropriate for use in a population with exclusively chronic LBP. However, given the relatively low values for the AUC (Table 3.4) and the location of the other potential cut-points on the ROC curves (Figure 3.3), using different cut-off scores is unlikely to have improved the tool's performance in the current study.

The natural course of acute and chronic LBP has been shown to differ (Costa et al., 2012, Dunn and Croft, 2006, Grotle et al., 2010, Hayden et al., 2010). Grotle and colleagues (2010) reported that disability improved from baseline to 12 month follow-up in both a group with acute/subacute LBP and a group with chronic LBP. However, the improvement in disability was smaller in the group with chronic LBP in comparison to the group with acute/subacute LBP (Grotle et al., 2010). Similarly, previous studies have reported that those with chronic LBP showed a more "persistent course" compared to those with acute LBP (Costa et al., 2009, Hayden et al., 2010). Hence, if a patient with chronic LBP typically

does not change in their presentation much over time, the predictive ability of the SBST will be relatively low. An alternative explanation relates to the potential fluctuating nature of LBP. Dunn and colleagues (2006) recruited a sample with variable duration LBP and identified four different clusters that characterised the course of LBP over a twelve month period. One of these clusters demonstrated a fluctuating course of pain and disability (Dunn et al., 2006). In the current study, the follow-up measures were only collected at one point in time and therefore may not have best captured the “true” outcome.

Chronic LBP is complex and multidimensional (Waddell, 2006). It may be that the dichotomised response option (agree or disagree) used in the SBST was not sensitive enough to accurately or fully capture the risk profile of patients with chronic LBP. Furthermore, there are a large number of factors that may influence prognosis prediction and future recovery that are unrelated to the content of the SBST. Although the brevity and simplicity of the SBST is arguably one of its strengths, it is not unexpected that a screening tool as brief and as simple as the SBST was not able to strongly predict or discriminate the follow-up measures of interest in a population with chronic LBP.

In summary, there are a number of possible explanations as to why the SBST’s performance differed depending on the follow-up measure evaluated and why the SBST’s performance was relatively weaker in an Australian population with chronic LBP in comparison to previous reports.

4.6 Strengths of the current study

The current study has a number of strengths. A large sample was recruited with a very low attrition rate at the one year follow-up. There were a low number of missing values for both the baseline and the follow-up measures. Strengths of the cross sectional study were that it collected data on pain, disability, and a diverse range psychological measures including three novel measures not previously investigated in relation to the SBST (perceived risk of pain persistence, self-efficacy, and chronic pain acceptance). Follow-up measures were multi-domain, patient relevant, and in line with recommendations for research involving participants with chronic pain (Dworkin et al., 2005). The recovery indices selected, particularly for disability, allowed for direct comparisons across multiple studies. The follow-up period of one year after entry into the study made this one of the few studies that has evaluated the predictive and discriminative ability of the SBST at a

follow-up time point greater than six months. A range of statistical methods were used to explore the SBST's predictive and discriminative ability, facilitating comparison with previous research.

4.7 Limitations of the current study

There are a number of limitations related to the current study. The participants in this study were self-selecting volunteers predominantly from the general community with 21.4% ($n=62$) of the study participants recruited from a specific clinical setting, mostly private physiotherapy clinics. This may limit the ability to generalise the results to specific clinical populations. However, as stated previously, this cohort was largely a care seeking group and therefore is broadly representative of those suffering from chronic LBP in a primary care setting. In order to elicit information on LBP duration, participants were asked to recall how long they had had their back pain for. This may have resulted in recall bias. The participants in the current study had a median duration of LBP of 10 years suggesting that the results may be most generalisable to a group with similar duration LBP. This study recruited participants with dominant axial LBP therefore the results are most able to be generalised to this group. To be eligible to participate in this study, a minimum level of pain and disability was required. Hence, the results may not be generalisable to those suffering from chronic LBP who have very low levels of pain or disability. As discussed previously, the cohort in the current study was not highly disabled and therefore the results may also not be generalisable to those suffering from chronic LBP who are more severely disabled. Despite the aforementioned, each of the SBST risk subgroups were well represented and each subgroup had clinically meaningful levels of pain and disability.

Although a number of important and novel psychological measures were collected at baseline, other important constructs such as coping strategies and social factors were not evaluated. Cut-off scores to indicate non-recovery at the one year follow-up may not have accurately represented all participants who viewed themselves as recovered or not recovered. This was evident when considering the variability in the proportion of participants deemed recovered versus not recovered for the three follow-up measures. This study dichotomised the follow-up measures, which makes the results easy for readers to understand and interpret but may have resulted in a loss of information. As an example, a participant sitting just above and another sitting just below the cut-off point on a follow-up measure would have had a very similar score but would have been categorised

differently. Furthermore, using different operational definitions for recovery/non-recovery may have influenced the results of this study. Other patient centered measures of recovery such as quality of life and psychological measures were also not investigated.

The participants were free to pursue any intervention or treatment they wanted to throughout the study period. This study did not collect robust data on the type of treatments participants received or how helpful they perceived any treatment to have been. Hence, it is unknown what influence, if any, treatment may have had on recovery or the performance of the SBST. Nonetheless, the SBST was designed to be used to guide decision-making in primary or first contact care settings regardless of what treatment the patient may or may not have in the future.

Finally, this study does not provide any information on the clinical or economic benefits of stratified care. However, if the SBST has relatively weak prognostic ability, its usefulness to stratify care may be limited in this group (Kent et al., 2015).

4.8 Implications for clinical practice

In light of the strengths and despite the limitations acknowledged above, the results of this study have a number of important implications for clinical practice, both for health care providers and health care systems. The cross sectional study demonstrated that the SBST risk subgroups were able to differentiate increasing levels of pain, disability, and negative psychological affect and cognitions however, the differences between all three risk subgroups may not have been clinically meaningful or relevant. Subgroup differences appeared to have been limited to a low and a high risk group, rather than a low, medium, and high risk group in this cohort with chronic LBP. The prospective study demonstrated that the SBST had moderate predictive and acceptable discriminative ability for disability at the one year follow-up. The SBST's ability to predict and discriminate future pain intensity was relatively weak and the tool was unable to identify those participants who perceived themselves as improved versus not improved at the one year follow-up. In considering the aforementioned, the SBST should be applied with caution in patients presenting with chronic LBP. In considering the results from the current study, the implications for clinical practice are as follows:

- The SBST has value as a substitute for the administration of multiple full-length, unidimensional questionnaires for first-line screening to help identify patients

presenting with higher levels of pain, disability, and negative psychological affect and cognitions. Therefore the tool can alert the clinician to the need for further assessment.

- Clinicians may need to consider screening with alternative tools or use full-length questionnaires to elicit more detailed information to guide interventions.
- Reliance on the SBST as a sole indicator of prognosis in chronic axial LBP is not recommended. Although the SBST provided an acceptable indication of future disability, clinicians should be aware that patient relevant outcomes extend beyond disability. Therefore, the SBST may not capture a high risk of poor outcome from every patient's perspective.
- Although stratified care based on the SBST risk subgroups has shown promising results in other populations (Hill et al., 2011, Whitehurst et al., 2012), the design of health care systems or care pathways that stratify care based exclusively on the SBST in populations with chronic LBP in Australia cannot be recommended due to the relatively weak prognostic ability of the SBST in this population in comparison to a UK primary care setting where patients reported LBP of variable duration.

Clinician's need to consider the complex, multidimensional, and potentially fluctuating nature of chronic LBP. The brevity and simplicity of the SBST is one of its strengths but the tool may not adequately capture prognostic risk in more complex patient presentations. The SBST was designed to identify potentially modifiable prognostic indicators for poor outcome however, certain factors shown to demonstrate prognostic capability may not be modifiable through direct treatment such as age, sex, and pain episode duration. In addition, screening for risk of poor prognosis at a single point in time may not be adequate. Screening assessment over a number of time points may provide better prognostic information and more accurately capture the risk profile of the patient. This would also facilitate the implementation of targeted interventions that adapt to the patient's presentation over time. The SBST should not be used as a standalone tool but alongside the clinical examination and in conjunction with sound clinical reasoning, clinical intuition, and expert judgement.

4.9 Recommendations for future research

A number of avenues for future research are recommended. Two SBST risk subgroups rather than three may be more appropriate in different healthcare settings. A prospective

comparison of the predictive and discriminative ability of the SBST using two risk subgroups as opposed to three and the way in which this may translate into care pathways could be evaluated. Other authors have suggested the need to evaluate the appropriateness of two SBST risk categories to guide decision-making (Beneciuk et al., 2015).

This study demonstrated that the SBST risk subgroups were inversely related to the positive constructs of self-efficacy and chronic pain acceptance. Further research evaluating the role of adaptive coping and resiliency factors in predicting outcomes in chronic LBP is warranted. Further to this, evaluating whether or not the integration of these modifiable positive constructs into screening questionnaires is of value would be an interesting avenue of research. Targeted treatment strategies aimed at increasing self-efficacy and pain acceptance may be just as important as reducing the negative prognostic indicators such as anxiety, depression, and fear. Outcomes may be improved by facilitating resilience mechanisms which strengthen a person's ability to cope with pain (Sturgeon and Zautra, 2010). This area of research has been previously recommended (Wideman et al., 2012).

One of the limitations of the original development and validation of the SBST was that it was developed to predict future LBP related disability to the exclusion of other important measures such as pain, global perceived change/improvement, health related quality of life, psychological health, return to work/future sick leave, or healthcare utilisation. It has been suggested that defining recovery solely in terms of pain and/or disability may limit individual relevance (Kamper et al., 2010). Furthermore, standardised measures such as the RMDQ may not adequately capture the goals or interests of the patient (Nicholas and George, 2011). Studies investigating the ability of the SBST to predict the aforementioned outcomes are lacking as only future pain and global perceived change have received attention. Furthermore, a complex construct such as recovery is unlikely to be adequately or accurately captured by single item measures (Kamper et al., 2011). Therefore, research evaluating the ability of the SBST to predict or discriminate combined or composite measures of recovery is indicated. To date, only one study involving the SBST has used a composite measure of pain and disability as an outcome measure (George and Beneciuk, 2015). George and Beneciuk (2015) reported that the SBST was a robust predictor of this composite measure at the six month follow-up. Combined measures of global perceived change with pain and/or disability may also demonstrate better discriminatory ability between recovery and non-recovery (Mehling et al., 2011).

The prognostic value of repeated assessment with the SBST and the use of change scores has preliminary support (Beneciuck et al., 2014, Beneciuk et al., 2013, Wideman et al., 2012). Repeated assessment with the SBST over multiple time points may provide further prognostic information for patients with chronic LBP beyond that achieved by screening at initial intake only. However, further research is necessary to confirm this.

Finally, conflicting results are reported in the literature with respect to the value of using screening tools compared to using clinician's impression, judgement, and intuition in order to identify risk factors for poor prognosis or individuals at risk of a poor outcome. The majority of studies have compared clinicians and screening tools (Bishop and Foster, 2005, Calley et al., 2010, Dagfinrud et al., 2013, Daubs et al., 2010, Fersum et al., 2009, Grevitt et al., 1998, Haggman et al., 2004, Hill et al., 2010a, Jellema et al., 2007, Kongsted et al., 2015, Newell et al., 2013). A combination approach of the use of a screening questionnaire such as the SBST, clinical examination, and clinician judgement may be optimal (Beales et al., 2016) although pragmatic studies are lacking in this area. Research that investigates whether or not prognostic accuracy is improved if screening questionnaire results are integrated with clinical assessment would be of value from a clinical practice point of view.

4.10 Conclusions

The results of this study provide valuable information to clinicians on the usefulness and limitations of using the SBST in patients with chronic LBP in an Australian setting. The SBST should be applied with caution for patients presenting with chronic LBP. The SBST has value as a substitute for the administration of multiple full-length, unidimensional questionnaires to initially screen for high levels of pain, disability, and negative psychological affect and cognitions in clinical practice. Therefore the tool can alert the clinician to the need for further assessment. The SBST has moderate predictive and acceptable discriminative ability for future disability. However, clinicians need to be aware that patient relevant outcomes extend beyond disability and therefore the SBST may not capture a high risk of poor outcome from every patient's perspective. The SBST should be used alongside the clinical examination and in conjunction with sound clinical reasoning, clinical intuition, and expert judgement when making care decisions.

4.11 References

- BALAGUE, F., MANNION, A., PELLISE, F. & CEDRASCHI, C. 2012. Non-specific low back pain. *Lancet*, 379, 482-491.
- BEALES, D., KENDELL, M., CHANG, R., HÅMSØ, M., GREGORY, L., RICHARDSON, K. & O' SULLIVAN, P. 2016. Association between the 10 item Örebro Musculoskeletal Screening Questionnaire and physiotherapists' perception of the contribution of biopsychosocial factors in patients with musculoskeletal pain. *Manual Therapy*, 23, 48-55.
- BEATON, D., BOERS, M. & WELLS, G. 2002. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology*, 14, 109-114.
- BENECIUUK, J., FRITZ, J. & GEORGE, S. 2014. The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings. *Journal of Orthopaedic and Sports Physical Therapy*, 44, 656-664.
- BENECIUUK, J., BISHOP, M., FRITZ, J., ROBINSON, M., ASAL, N., NISENZON, A. & GEORGE, S. 2013. The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Physical Therapy*, 93, 321-333.
- BENECIUUK, J., ROBINSON, M. & GEORGE, S. 2015. Subgrouping for patients with low back pain: a multidimensional approach incorporating cluster analysis and the STarT Back Screening Tool. *The Journal of Pain*, 16, 19-30.
- BISHOP, A. & FOSTER, N. 2005. Do physical therapists in the United Kingdom recognize psychosocial factors in patients with acute low back pain? *Spine*, 30, 1316-1322.
- CALLEY, D., JACKSON, S., COLLINS, H. & GEORGE, S. 2010. Identifying patient fear-avoidance beliefs by physical therapists managing patients with low back pain. *Journal of Orthopaedic and Sports Physical Therapy*, 40, 774-783.
- COSTA, L., MAHER, C., HANCOCK, M., MCAULEY, J., HERBERT, R. & COSTA, L. 2012. The prognosis of acute and persistent low-back pain: a meta-analysis. *Canadian Medical Association Journal*, 184, E613-E624.
- COSTA, L., MAHER, C., MCAULEY, J., HANCOCK, M., HERBERT, R., REFSHAUGE, K. & HENSCHKE, N. 2009. Prognosis for patients with chronic low back pain: inception cohort study. *British Medical Journal*, 339, b3829.
- DAGFINRUD, H., STORHEIM, K., MAGNUSSEN, L., ODEGAARD, T., HOFTANISKA, I., LARSEN, L., RINGSTAD, P., HATLEBREKKE, F. & GROTE, M. 2013. The predictive validity of the Örebro Musculoskeletal Pain Questionnaire and the clinicians' prognostic assessment following manual therapy treatment of patients with LBP and neck pain. *Manual Therapy*, 18, 124-129.
- DAUBS, M., PATEL, A., WILICK, S., KENDALL, R., HANSEN, P., PETRON, D. & BRODKE, D. 2010. Clinical impression versus standardized questionnaire: the spinal surgeon's ability to assess psychological distress. *The Journal of Bone and Joint Surgery*, 92, 2878-2883.

- DIONNE, C., VON KORFF, M., KOEPESELL, T., DEYO, R., BARLOW, W. & CHECKOWAY, H. 2001. Formal education and back pain: a review. *Journal of Epidemiology and Community Health*, 55, 455-468.
- DUNN, K. & CROFT, P. 2006. The importance of symptom duration in determining prognosis. *Pain*, 121, 126-132.
- DUNN, K., JORDAN, K. & CROFT, P. 2006. Characterizing the course of low back pain: a latent class analysis. *American Journal of Epidemiology*, 163, 754-761.
- DWORKIN, R., TURK, D., FARRAR, J., HAYTHORNTHWAITE, J., JENSEN, M., KATZ, N., KERNS, R., STUCKI, G., ALLEN, R., BELLAMY, N., CARR, D., CHANDLER, J., COWAN, P., DIONNE, R., GALER, B., HERTZ, S., JADAD, A., KRAMER, L., MANNING, D., MARTIN, S., MCCORMICK, C., MCDERMOTT, M., MCGRATH, P., QUESSY, S., RAPPAPORT, B., ROBBINS, W., ROBINSON, J., ROTHMAN, M., ROYAL, M., SIMON, L., STAUFFER, J., STEIN, W., TOLLETT, J., WERNICKE, J. & WITTER, J. 2005. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, 113, 9-19.
- DWORKIN, R., TURK, D., WYRWICH, K., BEATON, D., CLEELAND, C., FARRAR, J., HAYTHORNTHWAITE, J., JENSEN, M., KERNS, R., ADER, D., BRANDENBURG, N., BURKE, L., CELLA, D., CHANDLER, J., COWAN, P., DIMITROVA, R., DIONNE, R., HERTZ, S., JADAD, A., KATZ, N., KEHLET, H., KRAMER, L., MANNING, D., MCCORMICK, C., MCDERMOTT, M., MCQUAY, H., PATEL, S., PORTER, L., QUESSY, S., RAPPAPORT, B., RAUSCHKOLB, C., REVICKI, D., ROTHMAN, M., SCHMADER, K., STACEY, B., STAUFFER, J., VON STEIN, T., WHITE, R., WITTER, J. & ZAVISIC, S. 2008. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *The Journal of Pain*, 9, 105-121.
- FERSUM, K., O'SULLIVAN, P., KVÅLE, A. & SKOUEN, J. 2009. Inter-examiner reliability of a classification system for patients with non-specific low back pain. *Manual Therapy*, 14, 555-561.
- FIELD, J. & NEWELL, D. 2012. Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropractic and Manual Therapies* 20, 17.
- FRITZ, J., BENECIUK, J., GEORGE, S., HILL, J. & HAY, E. 2011. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Physical Therapy*, 91, 722-736.
- GABEL, C., MELLOH, M., YELLAND, M., BURKETT, B. & ROIKO, A. 2011. Predictive ability of a modified Örebro Musculoskeletal Pain Questionnaire in an acute/subacute low back pain working population. *European Spine Journal*, 20, 449-457.
- GEORGE, S. & BENECIUK, J. 2015. Psychological predictors of recovery from low back pain: a prospective study. *BMC Musculoskeletal Disorders*, 16, 49.
- GLAS, A., LIJMER, J., PRINS, M., BONSEL, G. & BOSSUYT, P. 2003. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*, 56, 1129-1135.
- GREVITT, M., PANDE, K., O'DOWD, J. & WEBB, J. 1998. Do first impressions count? A comparison of subjective and psychologic assessment of spinal patients. *European Spine Journal*, 7, 218-223.

- GROTLE, M., FOSTER, N., DUNN, K. & CROFT, P. 2010. Are prognostic indicators for poor outcome different for acute and chronic low back pain consulters in primary care? *Pain*, 151, 790-797.
- GROTLE, M., VOLLESTAD, N. & BROX, J. 2006. Screening for yellow flags in first-time acute low back pain: reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. *The Clinical Journal of Pain*, 22, 458-467.
- HAGGMAN, S., MAHER, C. & REFSHAUGE, K. 2004. Screening for symptoms of depression by physical therapists managing low back pain. *Physical Therapy*, 84, 1157-1166.
- HAYDEN, J., CHOU, R., HOGG-JOHNSON, S. & BOMBARDIER, C. 2009. Systematic reviews of low back pain prognosis had variable methods and results - guidance for future prognostic reviews. *Journal of Clinical Epidemiology*, 62, 781-796.
- HAYDEN, J., DUNN, K., VAN DER WINDT, D. & SHAW, W. 2010. What is the prognosis of back pain? *Best Practice and Research Clinical Rheumatology*, 24, 167-179.
- HENEWEER, H., AUFDEMKAMPE, G., VAN TULDER, M., KIERS, H., STAPPAERTS, K. & VANHEES, L. 2007. Psychosocial variables in patients with (sub) acute low back pain. *Spine*, 32, 586-592.
- HENSCHKE, N., MAHER, C., REFSHAUGE, K., HERBERT, R., CUMMING, R., BLEASEL, J., YORK, J., DAS, A. & MCAULEY, J. 2008. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *British Medical Journal*, 337, a171.
- HILL, J., DUNN, K., LEWIS, M., MULLIS, R., MAIN, C., FOSTER, N. & HAY, E. 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632-641.
- HILL, J., DUNN, K., MAIN, C. & HAY, E. 2010b. Subgrouping low back pain: a comparison of the STarT Back Tool with the Örebro Musculoskeletal Pain Screening Questionnaire. *European Journal of Pain*, 14, 83-89.
- HILL, J., VOHORA, K., DUNN, K., MAIN, C. & HAY, E. 2010a. Comparing the STarT Back Screening Tool's subgroup allocation of individual patients with that of independent clinical experts. *The Clinical Journal of Pain*, 26, 783-787.
- HILL, J., WHITEHURST, D., LEWIS, M., BRYAN, S., DUNN, K., FOSTER, N., KONSTANTINOOU, K., MAIN, C., MASON, E., SOMERVILLE, S., SOWDEN, G., VOHORA, K. & HAY, E. 2011. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*, 378, 1560-1571.
- HOCKINGS, R., MCAULEY, J. & MAHER, C. 2008. A systematic review of the predictive ability of the Örebro Musculoskeletal Pain Questionnaire. *Spine*, 33, E494-E500.
- HOSMER, D., LEMESHOW, S. & STURDIVANT, R. 2013. *Wiley series in probability and statistics: applied logistic regression (3rd edition)*, New York, NY, USA, John Wiley & Sons.
- HUSH, J., REFSHAUGE, K., SULLIVAN, G., DE SOUZA, L., MAHER, C. & MCAULEY, J. 2009. Recovery: what does this mean to patients with low back pain? *Arthritis and Rheumatism*, 61, 124-131.

- JACKSON, T., WANG, Y., WANG, Y. & FAN, H. 2014. Self-efficacy and chronic pain outcomes: a meta-analytic review. *The Journal of Pain*, 15, 800-814.
- JELLEMA, P., VAN DER WINDT, D., VAN DER HORST, H., STALMAN, W. & BOUTER, L. 2007. Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. *British Journal of General Practice*, 57, 15-22.
- KAMPER, S., MAHER, C., HERBERT, R., HANCOCK, M., HUSH, J. & SMEETS, R. 2010. How little pain and disability do patients with low back pain have to experience to feel that they have recovered? *European Spine Journal*, 19, 1495-1501.
- KAMPER, S., STANTON, T., WILLIAMS, C., MAHER, C. & HUSH, J. 2011. How is recovery from low back pain measured? A systematic review of the literature. *European Spine Journal*, 20, 9-18.
- KENT, P., MIRKHIL, S., KEATING, J., BUCHBINDER, R., MANNICHE, C. & ALBERT, H. 2014. The concurrent validity of brief screening questions for anxiety, depression, social isolation, catastrophization, and fear of movement in people with low back pain. *The Clinical Journal of Pain*, 30, 479-489.
- KENT, P., MORSØ, L. & KONGSTED, A. 2015. Letter to the editor concerning 'Using the STarT Back Tool: does timing of stratification matter?'. *Manual Therapy*, 20, e13.
- KONGSTED, A., ANDERSEN, C., HANSEN, M. & HESTBAEK, L. 2015. Prediction of outcome in patients with low back pain - a prospective cohort study comparing clinicians' predictions with those of the Start Back Tool. *Manual Therapy*, 21, 120-127.
- KONGSTED, A., JOHANNESSEN, E. & LEBOEUF-YDE, C. 2011. Feasibility of the STarT Back Screening Tool in chiropractic clinics: a cross-sectional study of patients with low back pain. *Chiropractic and Manual Therapies*, 19, 10.
- KONSTANTINOOU, K., HIDER, S., JORDAN, J., LEWIS, M., DUNN, K. & HAY, E. 2013. The impact of low back-related leg pain on outcomes as compared with low back pain alone: a systematic review of the literature. *Clinical Journal of Pain*, 29, 644-654.
- LINTON, S. & BOERSMA, K. 2003. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. *The Clinical Journal of Pain*, 19, 80-86.
- LOVIBOND, S. & LOVIBOND, P. 1995. *Manual for the Depression Anxiety Stress Scales*, Sydney, Psychology Foundation.
- LUNDBERG, M., GRIMBY-EKMAN, A., VERBUNT, J. & SIMMONDS, M. J. 2011. Pain-related fear: a critical review of the related measures. *Pain Research and Treatment*, 2011, 494196.
- MCCRACKEN, L. & ECCLESTON, C. 2006. A comparison of the relative utility of coping and acceptance-based measures in a sample of chronic pain sufferers. *European Journal of Pain*, 10, 23-29.
- MEHLING, W., AVINS, A., ACREE, M., CAREY, T. & HECHT, F. 2015. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19, 439-446.

- MEHLING, W., GOPISETTY, V., ACREE, M., PRESSMAN, A., CAREY, T., GOLDBERG, H., HECHT, F. & AVINS, A. 2011. Acute low back pain and primary care: how to define recovery and chronification? *Spine*, 36, 2316-2323.
- MORSØ, L., ALBERT, H., KENT, P., MANNICHE, C. & HILL, J. 2011. Translation and discriminative validation of the STarT Back Screening Tool into Danish. *European Spine Journal*, 20, 2166-2173.
- MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.
- MORSØ, L., KENT, P., ALBERT, H. & MANNICHE, C. 2013a. Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care? *Manual Therapy*, 18, 54-59.
- MORSØ, L., KENT, P., MANNICHE, C. & ALBERT, H. 2014. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine Journal*, 23, 120-128.
- MORSØ, L., KONGSTED, A., HESTBAEK, L. & KENT, P. 2016. The prognostic ability of the STarT Back Tool was affected by episode duration. *European Spine Journal*, 25, 936-944.
- NEWELL, D., FIELD, J. & POLLARD, D. 2015. Using the STarT Back Tool: does timing of stratification matter? *Manual Therapy*, 20, 533-539.
- NEWELL, D., FIELD, J. & VISNES, N. 2013. Prognostic accuracy of clinicians for back, neck and shoulder patients in routine practice. *Chiropractic and Manual Therapies*, 21, 42.
- NICHOLAS, M., ASGHARI, A. & BLYTH, F. 2008. What do the numbers mean? Normative data in chronic pain measures. *Pain*, 134, 158-173.
- NICHOLAS, M. & GEORGE, S. 2011. Psychologically informed interventions for low back pain: an update for physical therapists. *Physical Therapy*, 91, 765-776.
- NONCLERCQ, O. & BERQUIN, A. 2012. Predicting chronicity in acute back pain: validation of a French translation of the Örebro Musculoskeletal Pain Screening Questionnaire. *Annals of Physical and Rehabilitation Medicine*, 55, 263-278.
- OSTELO, R., DEYO, R., STRATFORD, P., WADDELL, G., CROFT, P., VON KORFF, M., BOUTER, L. & DE VET, H. 2008. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine*, 33, 90-94.
- PAGE, I., ABBOUD, J., O. SHAUGHNESSY, J., LAURENCELLE, L. & DESCARREAU, M. 2015. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskeletal Disorders*, 16, 201.
- RAINVILLE, J., SOBEL, J., HARTIGAN, C. & WRIGHT, A. 1997. The effect of compensation involvement on the reporting of pain and disability by patients referred for rehabilitation of chronic low back pain. *Spine*, 22, 2016-2024.

- RAMOND, A., BOUTON, C., RICHARD, I., ROQUELAURE, Y., BAUFRETON, C., LEGRAND, E. & HUEZ, J. 2011. Psychosocial risk factors for chronic low back pain in primary care-a systematic review. *Family Practice*, 28, 12-21.
- SHAW, W., MAIN, C. & JOHNSTON, V. 2011. Addressing occupational factors in the management of low back pain: implications for physical therapist practice. *Physical Therapy*, 91, 777-789.
- SHAW, W., PRANSKY, G. & FITZGERALD, T. 2001. Early prognosis for low back disability: intervention strategies for health care providers. *Disability and Rehabilitation*, 23, 815-828.
- STURGEON, J. & ZAUTRA, A. 2010. Resilience: a new paradigm for adaptation to chronic pain. *Current Pain and Headache Reports*, 14, 105-112.
- SULLIVAN, M. 2009. *The Pain Catastrophizing Scale user manual* [Online]. Available: http://sullivan-painresearch.mcgill.ca/pdf/pcs/PCSTManual_English.pdf [Accessed 7 November 2015].
- TERWEE, C., ROORDA, L., DEKKER, J., BIERMA-ZEINSTRAS, S., PEAT, G., JORDAN, K., CROFT, P. & DE VET, H. 2010. Mind the MIC: large variation among populations and methods. *Journal of Clinical Epidemiology*, 63, 524-534.
- TRAEGER, A., HENSCHKE, N., HUBSCHER, M., WILLIAMS, C., KAMPER, S., MAHER, C., MOSELEY, G. & MCAULEY, J. 2015. Development and validation of a screening tool to predict the risk of chronic low back pain in patients presenting with acute low back pain: a study protocol. *BMJ Open*, 5, e007916.
- VON KORFF, M., SHORTREED, S., SAUNDERS, K., LERESCHE, L., BERLIN, J., STANG, P. & TURNER, J. 2014. Comparison of back pain prognostic risk stratification item sets. *The Journal of Pain*, 15, 81-89.
- WADDELL, G. 2006. Preventing incapacity in people with musculoskeletal disorders. *British Medical Bulletin*, 77 and 78, 55-69.
- WHITEHURST, D., BRYAN, S., LEWIS, M., HILL, J. & HAY, E. 2012. Exploring the cost-utility of stratified primary care management for low back pain compared with current best practice within risk-defined subgroups. *Annals of the Rheumatic Diseases*, 71, 1796-1802.
- WIDEMAN, T., HILL, J., MAIN, C., LEWIS, M., SULLIVAN, M. & HAY, E. 2012. Comparing the responsiveness of a brief, multidimensional risk screening tool for back pain to its unidimensional reference standards: the whole is greater than the sum of its parts. *Pain*, 153, 2182-2191.
- WRIGHT, A., HANNON, J., HEGEDUS, E. & KAVCHAK, A. 2012. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *Journal of Manual and Manipulative Therapy*, 20, 160-166.
- WRIGHT, M., WREN, A., SOMERS, T., GOETZ, M., FRAS, A., HUH, B., ROGERS, L. & KEEFFE, F. 2011. Pain acceptance, hope, and optimism: relationships to pain and adjustment in patients with chronic musculoskeletal pain. *The Journal of Pain*, 12, 1155-1162.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Appendix A: Ethics Approval



Memorandum

To	Michelle Kendell
From	A/Prof. Helen Slater
Subject	Protocol Approval PT262/2013
Date	18 December 2013
Copy	Darren Beales. Martin Rabey

Office of Research and Development

Human Research Ethics Committee

Telephone 9266 2784

Facsimile 9266 3793

Email hrec@curtin.edu.au

Thank you for your Application for Approval of Research with Low Risk (Ethical Requirements) for the project titled "Predicting Outcome with a Short Musculoskeletal Pain Screening Questionnaire: A Prospective Study".

Please make the following minor amendments based on reviewers' comments:

The participant information sheet should be amended: under 'Benefits' the word 'telling' should be amended to 'tell'

On behalf of the Human Research Ethics Committee, I am authorised to inform you that this project is approved.

Your approval has the following conditions.

- (i) An annual progress report/completion report (attached) on the project must be submitted to the School of Physiotherapy, Ethics Coordinator.
- (ii) It is your responsibility, as the researcher, to meet the conditions outlined above and to retain the necessary records demonstrating that these have been completed.

Approval of this project is for a period of four years from the **December 18th 2013 to December 18th 2017**.

The approval number for your project is **PT262/2013**. Please quote this number in any future correspondence. If at any time during the approval term changes/amendments occur, or if a serious or unexpected adverse event occurs, please advise me immediately.

A handwritten signature in blue ink, appearing to read "Helen Slater", with a long horizontal flourish extending to the right.

Associate Professor Helen Slater
Coordinator, Ethics Committee
School of Physiotherapy and Exercise Science
Curtin University

Appendix B: Participant Information Sheet

Note: Baseline data for the current study was obtained from this study (see Methods, Chapter2)



Faculty of Health Sciences
School of Physiotherapy and Exercise Science

Telephone +61 8 9266 4644

Facsimile +61 8 9266 3699

Web www.physiotherapy.curtin.edu.au

PARTICIPANT INFORMATION SHEET

PLAIN LANGUAGE VERSION

Low Back Pain: Identifying Factors Associated With Persistent Low Back Pain

Principal Investigator:	Martin Rabey
Project Supervisor:	Professor Peter O'Sullivan
Co-Supervisors:	Dr. Darren Beales, Dr. Anne Smith, Associate Professor Helen Slater

Purpose of Research: Back pain can come from strained muscles or sprained joints. These usually heal and then the pain gets better. However, in some people back pain can carry on. This is called persistent or chronic back pain. It is a difficult problem and why some people have ongoing pain is not clear. There is not a strong link between what we see on x-rays or scans and the pain people feel. People with persistent pain often have other problems like low mood, fear of moving in certain ways, or worries about work.

This research is to try to increase what we know about the complex mix of factors involved in persistent back pain. We need to study people with persistent low back pain and see how these factors interact. If we study a lot of people with back pain, we may be able to identify different groups. We will measure many different factors in 300 people and then look at this information. We are also interested in any treatment people have for their back pain in the year after you join the study. So one year later we will contact you to see if your pain has changed, and find out what treatments you had. You can have any treatment you like during this time.

Your role: We would like you to be a participant in our research. We need to measure different factors to do with your pain. The study has 3 parts:
Part 1 involves you completing some questions at home. It takes 30-40 minutes.
Part 2 involves a visit to Curtin University for some tests of different sensations (pressure, touch, heat and cold). This takes 60-90 minutes.
Part 3 involves completing some more questions one year later. This takes 10 minutes.

Part 1: We will give you some questions to complete in your own time. They ask about your pain, how it makes you feel, and how it affects your work and hobbies. If you lose concentration you can take a break and finish the questions later. Some questions ask if you feel "low or down" or anxious. Depending on your response, we might send a letter to your doctor to make sure you get the best care for your pain problem. We will inform you if we send a letter.

Part 2: We will arrange for you to come to the School of Physiotherapy at Curtin University, Bentley. We will conduct some sensory (e.g.; touch, pressure) tests with you. We will ask you to lie on your tummy to test these sensations on your back and your wrist. We test the wrist because you have no pain there.

These are the sensation tests:

- 1) Touch: We lightly touch the skin in one place or two places. We ask you how many places you felt.
- 2) Touch: Using plastic “hairs” we touch your skin and ask if you can feel it. We will also repeatedly touch your skin with the “hairs” and see if you feel any pain.
- 3) Pressure: Using a pencil, a make-up brush, a toothpick, and a tuning fork, we will see if you can feel the pencil, the brushing, the toothpick and the fork’s vibrations. We will ask whether these cause any pain.
- 4) Pressure: Using a special device we will find the point at which pressure changes to pain. The test stops the moment you feel pain.
- 5) Heat/cold: Using another device we will find the point at which hot and cold becomes painful. Each test stops the moment you feel pain.
- 6) Body’s own pain control: Tests 4 and 5 are combined to see how effective your body’s own pain control is. We will put the warm probe on your wrist. Then we measure how much discomfort you feel when we put pressure to your back. This lasts about one minute.
- 7) We will then ask you to lie still and focus your attention on your back pain for one minute. We will measure if focusing changes your pain.

You may feel some discomfort during these last two tests. This should not last long. You can stop any time you wish.

Finally we will ask you to do some movements. You will be asked to try 20 forward bends and 20 backward bends. If your pain gets too strong you may stop at any time. During the movements video cameras will record how you move. We will ask you if your pain changes as you move.

Part 3: One year later we will send you a few questions in the mail. You complete them and mail them back to us. The questions ask about your pain and the treatments you have had.

Risks and Discomforts: You may feel some discomfort or pain during some tests. This is similar to what you might feel if you went to see your doctor or physiotherapist for your back pain, and they asked you to move your back and pressed to see where it hurts. Any increase in pain should settle down quickly once we have finished. If any tests are too painful you can tell the researcher and stop the tests. We will ask you some questions about your mood or your feelings.

Benefits: After your visit to the School of Physiotherapy we will tell you about your results. When we have all of the results we will tell you about the outcomes. We will hold a talk, which you can attend, where we discuss the results. The aim of the research is to improve our understanding of back pain. This may help find better treatments for back pain in time.

Confidentiality: We keep your details confidential by giving you an identification number. Your name will not be on any questionnaires or record sheets. It will only be on

the consent form. All forms and video tapes will be stored in a locked cabinet in the School of Physiotherapy. Information on computer will be protected by a password. Only the researchers will have access to the information. Information will be kept for seven years after we have published the results.

Refusal or Withdrawal: We ask you to sign a consent form. This says that you agree to take part in the research. Whether you take part in the research or not is up to you. You may refuse to ever be involved in the research. You may withdraw your consent whenever you like, without prejudice. Please ask any questions before signing the consent form.

Further Information: If you would like more information please contact Martin Rabey at: School of Physiotherapy, Curtin University, Kent St., Bentley, Western Australia 6102; Tel: 0487007116; e-mail: martin.rabey@postgrad.curtin.edu.au.

This study has been approved by the Curtin University Human Research Ethics Committee (Approval Number HR112/2012). The Committee is composed of members of the public, academics, lawyers, doctors and pastoral carers. If needed, verification of approval can be obtained either by writing to the Curtin University Human Research Ethics Committee, c/- Office of Research and Development, Curtin University, GPO Box U1987, Perth 6845, or by telephoning 9266 9223 or by e-mailing hrec@curtin.edu.au.

Appendix C: Written Informed Consent Sheet

Note: Baseline data for the current study was obtained from this study (see Methods, Chapter2)



Faculty of Health Sciences
School of Physiotherapy and Exercise Science

Telephone +61 8 9266 4644

Facsimile +61 8 9266 3699

Web www.physiotherapy.curtin.edu.au

CONSENT SHEET

This study has been approved by the Curtin University Human Research Ethics Committee (Approval Number HR112/2012).

- I understand the purpose and procedures of the study.
- I have been provided with the participant information sheet.
- I understand that the procedure itself may not benefit me.
- I understand that my involvement is voluntary and I can withdraw at any time without problem.
- I understand that a letter will be sent to my family doctor if I score highly on the questionnaire concerned with anxiety or depression.
- I understand that no personal identifying information like my name and address will be used and that all information will be securely stored for 7 years before being destroyed.
- I have been given the opportunity to ask questions.
- I agree to participate in the study outlined to me.

Signature

Date

Witness Signature

Date

Appendix D: Discussion Supplement - Subgroup Proportions

Summary of Baseline STarT Back Screening Tool Risk Subgroup Proportions

Study	STarT Back Screening Tool risk subgroup (%)		
	Low	Medium	High
Current study	28	40	32
Studies in chronic LBP			
Morsø et al., 2014	28	33	40
Page et al., 2015	66	21	13
Studies in LBP of variable duration			
Hill et al., 2008			
Developmental sample	40	34	25
External sample	47	37	15
Hill et al., 2011	26	46	28
Fritz et al., 2011	33	48	19
Kongsted et al., 2011	59	29	11
Morsø et al., 2011	40	34	26
Field and Newell, 2012	42	32	27
Morsø et al., 2013b	38	39	23
Beneciuk et al., 2013, Beneciuk et al., 2015, George and Beneciuk, 2015	36	38	26
Newell et al., 2015	39	37	24
Studies in acute LBP			
Mehling et al., 2015	32	46	22

LBP = low back pain.

References

- BENECIUK, J., BISHOP, M., FRITZ, J., ROBINSON, M., ASAL, N., NISENZON, A. & GEORGE, S. 2013. The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Physical Therapy*, 93, 321-333.
- BENECIUK, J., ROBINSON, M. & GEORGE, S. 2015. Subgrouping for patients with low back pain: a multidimensional approach incorporating cluster analysis and the STarT Back Screening Tool. *The Journal of Pain*, 16, 19-30.

- FIELD, J. & NEWELL, D. 2012. Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropractic and Manual Therapies* 20, 17.
- FRITZ, J., BENECIUK, J., GEORGE, S., HILL, J. & HAY, E. 2011. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Physical Therapy*, 91, 722-736.
- GEORGE, S. & BENECIUK, J. 2015. Psychological predictors of recovery from low back pain: a prospective study. *BMC Musculoskeletal Disorders*, 16, 49.
- HILL, J., DUNN, K., LEWIS, M., MULLIS, R., MAIN, C., FOSTER, N. & HAY, E. 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632-641.
- HILL, J., WHITEHURST, D., LEWIS, M., BRYAN, S., DUNN, K., FOSTER, N., KONSTANTINOOU, K., MAIN, C., MASON, E., SOMERVILLE, S., SOWDEN, G., VOHORA, K. & HAY, E. 2011. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*, 378, 1560-1571.
- KONGSTED, A., JOHANNESSEN, E. & LEBOEUF-YDE, C. 2011. Feasibility of the STarT Back Screening Tool in chiropractic clinics: a cross-sectional study of patients with low back pain. *Chiropractic and Manual Therapies*, 19, 10.
- MEHLING, W., AVINS, A., ACREE, M., CAREY, T. & HECHT, F. 2015. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19, 439-446.
- MORSØ, L., ALBERT, H., KENT, P., MANNICHE, C. & HILL, J. 2011. Translation and discriminative validation of the STarT Back Screening Tool into Danish. *European Spine Journal*, 20, 2166-2173.
- MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.
- MORSØ, L., KENT, P., MANNICHE, C. & ALBERT, H. 2014. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine Journal*, 23, 120-128.
- NEWELL, D., FIELD, J. & POLLARD, D. 2015. Using the STarT Back Tool: does timing of stratification matter? *Manual Therapy*, 20, 533-539.
- PAGE, I., ABBOUD, J., O. SHAUGHNESSY, J., LAURENCELLE, L. & DESCARREAU, M. 2015. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskeletal Disorders*, 16, 201.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Appendix E: Discussion Supplement - Follow-up Disability

Proportion of Participants Disabled at Follow-up (RMDQ ≥ 7) – Total Cohort and Stratified by Baseline STarT Back Screening Tool Risk Subgroup

Study	LBP duration	Follow-up period	Total cohort (%)	STarT Back Screening Tool risk subgroup (%)		
				Low	Medium	High
Current study	Chronic	1 year	31	15	35	43
Hill et al., 2008 UK primary care*	Variable	6 months	40	17	53	78
Morsø et al., 2013b Danish primary care	Variable	3 months	47	24	57	64
Morsø et al., 2013b UK primary care	Variable	3 months	36	17	54	78
Morsø et al., 2014 Danish secondary care	≈ 80% Chronic	6 months	68	48	71	80
Morsø et al., 2014 Danish primary care†	Variable	6 months	40	20	47	69
Mehling et al., 2015‡ USA primary care	Acute	6 months	22	15	27	13
Mehling et al., 2015‡ USA primary care	Acute	2 years	25	15	31	15

RMDQ = Roland Morris Disability Questionnaire; LBP = low back pain.

* External sample.

† Participants recruited from physiotherapy clinics.

‡ A modified item set of the STarT Back Screening Tool used.

References

- HILL, J., DUNN, K., LEWIS, M., MULLIS, R., MAIN, C., FOSTER, N. & HAY, E. 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632-641.
- MEHLING, W., AVINS, A., ACREE, M., CAREY, T. & HECHT, F. 2015. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19, 439-446.

MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.

MORSØ, L., KENT, P., MANNICHE, C. & ALBERT, H. 2014. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine Journal*, 23, 120-128.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Appendix F: Discussion Supplement - Disability Risk Ratios

Predictive ability of the STarT Back Screening Tool for Disability (RMDQ ≥ 7) using Risk Ratios

Study	LBP duration	Follow-up period	STarT Back Screening Tool risk subgroup		
			Low*	Medium RR [95% CI]	High RR [95% CI]
Current study	Chronic	1 year	REF	2.3 [1.3, 4.1]	2.9 [1.6, 5.1]
Morsø et al., 2013b Danish primary care	Variable	3 months	REF	2.4 [1.7, 3.4]	2.7 [1.8, 3.8]
Morsø et al., 2013b UK primary care	Variable	3 months	REF	3.1 [2.5, 3.9]	4.5 [3.6, 5.6]
Morsø et al., 2014 Danish secondary care	≈ 80% Chronic	6 months	REF	1.5 [1.3, 1.7]	1.7 [1.5, 2.0]
Morsø et al., 2014 Danish primary care†	Variable	6 months	REF	2.3 [1.2, 4.5]	3.5 [1.8, 6.6]

RMDQ = Roland Morris Disability Questionnaire; LBP = low back pain; RR = risk ratio; 95% CI = 95% confidence interval; REF = reference.

* Low risk subgroup used as the reference category.

† Participants recruited from physiotherapy clinics.

References

- MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.
- MORSØ, L., KENT, P., MANNICHE, C. & ALBERT, H. 2014. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine Journal*, 23, 120-128.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Appendix G: Discussion Supplement - Disability Area Under the Curve

Discriminative Ability of the STarT Back Screening Tool Baseline Total Score for Disability (RMDQ ≥ 7) using the AUC

Study	LBP duration	Follow-up period	AUC [95% CI]
Current study	Chronic	1 year	0.71 [0.64, 0.77]
Morsø et al., 2013b Danish primary care	Variable	3 months	0.71 [0.66, 0.77]
Morsø et al., 2013b UK primary care	Variable	3 months	0.81 [0.78, 0.84]
Morsø et al., 2014 Danish secondary care	≈ 80% Chronic	6 months	0.69 [0.66, 0.73]
Morsø et al., 2014 Danish primary care*	Variable	6 months	0.73 [0.64, 0.82]
Page et al., 2015† Canada general community	Chronic	6 months	0.84 [0.69, 1.00]
Page et al., 2015† Canada general community	Chronic	1 year	0.82 [0.61, 1.00]
Mehling et al., 2015‡ USA primary care	Acute	6 months	Range 0.54-0.63§
Mehling et al., 2015‡ USA primary care	Acute	2 years	Range 0.54-0.63§

RMDQ = Roland Morris Disability Questionnaire; LBP = low back pain; AUC = area under the curve; 95% CI = 95% confidence interval.

* Participants recruited from physiotherapy clinics.

† Oswestry Disability Questionnaire $\geq 24\%$ which equates to RMDQ ≥ 7 .

‡ A modified item set of the STarT Back Screening Tool used.

§ Specific values for the AUC were not reported for each of the outcome measures. The AUC for disability falls somewhere within this range.

References

MEHLING, W., AVINS, A., ACREE, M., CAREY, T. & HECHT, F. 2015. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19, 439-446.

- MORSØ, L., KENT, P., ALBERT, H., HILL, J., KONGSTED, A. & MANNICHE, C. 2013b. The predictive and external validity of the STarT Back Tool in Danish primary care. *European Spine Journal*, 22, 1859-1867.
- MORSØ, L., KENT, P., MANNICHE, C. & ALBERT, H. 2014. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine Journal*, 23, 120-128.
- PAGE, I., ABBOUD, J., O. SHAUGHNESSY, J., LAURENCELLE, L. & DESCARREAUX, M. 2015. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskeletal Disorders*, 16, 201.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Appendix H: Discussion Supplement - Disability Likelihood

Ratios

Discriminative Ability of the STarT Back Screening Tool Risk Subgroups for Disability (RMDQ ≥ 7) - Likelihood Ratios, Sensitivity, Specificity, and the Diagnostic Odds Ratio

Study	LBP duration	Follow-up period	LR+ [95% CI]	LR [95% CI]	Sens (%)	Spec (%)	DOR
Current study	Chronic	1 year					
L vs. M/H			1.36 [1.18, 1.57]	0.39 [0.22, 0.68]	85.5	37.0	3.48
L/M vs. H			1.67 [1.15, 2.43]	0.79 [0.65, 0.96]	39.8	76.2	2.12
Hill et al., 2008	Variable	6 months					
L vs. M/H			2.32 [1.96, 2.76]	0.30 [0.23, 0.40]	80.1	65.4	7.73
L/M vs. H			5.51 [3.30, 9.28]	0.74 [0.67, 0.81]	29.6*	94.6	7.45
Page et al., 2015†	Chronic	6 months					
L vs. M/H			$\leq 2.96^\ddagger$	$\geq 0.35^\ddagger$	$\approx 68^\ddagger$	$\approx 79^\ddagger$	$\approx 8^\ddagger$
L/M vs. H			–	–	–	–	–
Page et al., 2015†	Chronic	1 year					
L vs. M/H			$\leq 2.96^\ddagger$	$\geq 0.35^\ddagger$	$\approx 77^\ddagger$	$\approx 75^\ddagger$	$\approx 10^\ddagger$
L/M vs. H			–	–	–	–	–
Mehling et al., 2015§	Acute	6 months					
L vs. M/H			1.31	0.48	81.7	32.1	2.73
L/M vs. H			1.61	0.85	32.1	80.1	1.89
Mehling et al., 2015§	Acute	2 years					
L vs. M/H			1.27	0.58	77.1	39.4	2.19
L/M vs. H			1.63	0.82	36.6	77.6	1.99

Likelihood ratio confidence intervals are included where available. RMDQ = Roland Morris Disability Questionnaire; LBP = low back pain; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; Sens = sensitivity; Spec = specificity; DOR = diagnostic odds ratio; 95% CI = 95% confidence interval; L = low risk group; M = medium risk group; H = high risk group; – = data not available.

* As noted by Mehling and colleagues (2015), the value of 39.6% in Table 4 in Hill and colleagues (2008) appears incorrect. Sensitivity was recalculated as 58/196 (= 29.6%) based on the numbers provided in the text on page 638 in Hill and colleagues (2008).

† Oswestry Disability Questionnaire $\geq 24\%$ which equates to RMDQ ≥ 7 .

‡ Specific values for the likelihood ratios were not reported. Sensitivity and specificity values were not reported in the text and are therefore estimated from Figure 5. The DOR was calculated from the estimated sensitivity and specificity values.

§ A modified item set of the STarT Back Screening Tool used.

References

- HILL, J., DUNN, K., LEWIS, M., MULLIS, R., MAIN, C., FOSTER, N. & HAY, E. 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632-641.
- MEHLING, W., AVINS, A., ACREE, M., CAREY, T. & HECHT, F. 2015. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19, 439-446.
- PAGE, I., ABBOUD, J., O. SHAUGHNESSY, J., LAURENCELLE, L. & DESCARREAU, M. 2015. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskeletal Disorders*, 16, 201.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.