

**CONTEMPORARY MANAGEMENT
OF LOW BACK PAIN**

Leonardo Oliveira Pena Costa, PT, MSc

Thesis presented for the degree of

Doctor of Philosophy

The University of Sydney

2009

Supervisor's Statement

As supervisors of Leonardo Oliveira Pena Costa's doctoral work, we certify that we consider his thesis "Contemporary Management of Low Back Pain" to be suitable for examination.

Professor Christopher G. Maher

Faculty of Medicine,

The University of Sydney

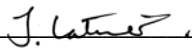


Date 20.03.2009

Associate Professor Jane Latimer

Faculty of Medicine,

The University of Sydney



Date 20.03.2009

Candidate's Statement

I, Leonardo Oliveira Pena Costa, hereby declare that this submission is my own work and that it contains no material previously published or written by another person except where acknowledged in the text. Nor does it contain material which has been accepted for the award of another degree.

I, Leonardo Oliveira Pena Costa, understand that if I am awarded a higher degree for my thesis entitled "Contemporary Management of Low Back Pain" being lodged herewith for examination, the thesis will be lodged in the University Library and be available immediately for use. I agree that the University Librarian (or in the case of a department, the Head of the Department) may supply a photocopy or microform of the thesis to an individual for research or study or to a library.

Signed _____

Date 20.03.2009

Table of Contents

	PAGE
Supervisor's Statement	ii
Candidate's Statement.....	iii
Table of Contents	iv
Acknowledgements	ix
Publications and Presentations.....	x
Preface.....	xiii
Abstract	xv
Chapter One: Introduction	1
1.1 Introduction to low back pain	2
1.2 Economic burden of low back pain.....	2
1.3 Low back pain classification.....	2
1.4 The prognosis of low back pain	4
1.5 Treatment possibilities for low back pain	5
1.5.1 Exercise as a treatment for chronic low back pain.....	6
1.5.2 Motor control exercise for chronic low back pain	8
1.5.3 Efficacy of motor control exercise for chronic low back pain.....	10
1.6 The assessment of motor control changes.....	11
1.6.1 Rehabilitative Ultrasound Imaging	11
1.6.2 Reproducibility of Rehabilitative Ultrasound Imaging.....	12
1.7 Self-report outcome measures for low back pain.....	13
1.7.1 Availability of self-report measures for low back pain.....	13
1.7.2 Cross-cultural adaptation and clinimetric studies in Brazil	15
1.8 Aims of the thesis.....	15
1.9 References	17

Chapter Two: The effect of motor control exercise versus placebo in patients with chronic low back pain	24
Statement of authorship contribution	25
Abstract	26
Background	26
Methods	29
Conclusion	32
References	32
Chapter Three: Exercise for chronic low back pain: A randomised, placebo-controlled trial	34
Statement of authorship contribution	35
Abstract	37
Introduction	38
Methods	38
Results	45
Discussion	47
References	52
Table 1	57
Table 2	58
Table 3	59
Figure 1	60
Figure 2	61
Chapter Four: Reproducibility of real-time ultrasound for the measurement of abdominal muscle activity: a systematic review	62
Statement of authorship contribution	63
Abstract	65

Introduction.....	66
Methods.....	68
Results.....	70
Discussion.....	72
References.....	76
Table 1.....	80
Table 2.....	81
Table 3.....	82
Table 4.....	83
Table 5.....	87
Table 6.....	91
Figure 1.....	92
Chapter Five: An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non- specific low back pain	93
Statement of authorship contribution	94
Abstract.....	96
Introduction.....	97
Methods.....	99
Results.....	103
Discussion.....	103
References.....	108
Table 1.....	111
Table 2.....	112
Figure 1.....	113
Figure 2.....	114

Chapter Six: The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments	115
Statement of authorship contribution.....	116
Abstract.....	117
Introduction.....	118
How to select a cross-cultural adaptation.....	120
Available Brazilian-Portuguese cross-cultural adaptations, an example from low back pain questionnaires	123
References.....	124
Chapter Seven: Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations	125
Statement of authorship contribution.....	126
Abstract.....	127
Introduction.....	127
Materials and Methods.....	127
Results.....	128
Discussion.....	130
References.....	134
Appendix 1: Quality criteria for measurement properties of health status questionnaires.....	137
Chapter Eight: Psychometric characteristics of the Brazilian-Portuguese versions of the Functional Rating Index and the Roland Morris Disability Questionnaire	138
Statement of authorship contribution.....	139
Abstract.....	140
Introduction.....	140
Methods.....	141
Results.....	142

Discussion.....	142
References.....	144
Appendix 1: Brazilian-Portuguese version of the Functional Rating Index	146
Chapter Nine: Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil. Which one is the best?	148
Statement of authorship contribution	149
Abstract.....	150
Introduction.....	150
Methods.....	151
Results.....	152
Discussion.....	153
References.....	154
Chapter Ten: Conclusions.....	155
10.1 Principal findings.....	156
10.2 Implications and suggestions for future research.....	157
10.2.1 Motor control exercise for chronic low back pain	157
10.2.2 Reproducibility of Rehabilitative Ultrasound Imaging measures of abdominal wall muscle activity.....	161
10.2.3 Cross-cultural adaptation and clinimetric testing of low back pain self-report outcome measures	162
10.3 References.....	164
Appendices.....	167
Appendix A: Guidelines for publication for the journal <i>PLoS Medicine</i>	168
Appendix B: Guidelines for publication for the journal <i>Physical Therapy</i>	180
Appendix C: Guidelines for publication for the journal <i>European Spine Journal</i>	196

Acknowledgements

I have heard that a simple decision can change our lives; I think that I made the best decision of my academic career when I chose Chris Maher and Jane Latimer to be my PhD supervisors. Your competence, knowledge, patience and friendship were key elements to making my PhD such an amazing and enjoyable experience. From the bottom of my heart I would like to thank you. You are simply the best!

I would like to show my appreciation to CAPES – Education Ministry / Brazil and the Catholic University of Minas Gerais/Brazil for providing me with financial support over the course of my PhD.

I would also like to thank my friends and staff from the University of Sydney, The George Institute for International Health and Liverpool Hospital. In particular I would like to thank Paulo Ferreira, Manuela Ferreira, Mark Hancock, Rob Herbert and my two “brothers” Marcos de Noronha and Steve Kamper for all of their support and friendship.

To my family in Brazil, people that I love so much, thank you for always being around me. A special thanks to my lovely mum, my sister Núbia & family, my uncle Ronaldo and my ‘in-law’ family. Dad, I miss you and I think about you every single day, I wish you were here to share this moment with me.

During these four years living in Sydney I was blessed to have the company of the most incredible woman I have ever known. I consider myself the luckiest person in the world to have Luciola as my wife. Lu, thanks for your love and for the support that you have given to me. You are the biggest source of encouragement that I have ever had. Without you this dream would not have been achievable. This thesis is fully dedicated to you.

Publications and Presentations

Much of the work presented in this thesis has been published and/or presented in the following forums:

Publications

Costa LOP, Maher CG, Latimer J (2006) Letters. *Spine* 31:2405-2407.

Costa LOP, Maher CG, Latimer J (2007) Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations. *Spine* 32:1028-1037.

Costa LOP, Maher CG, Latimer J, Ferreira PH, Pozzi GC, Ribeiro RN (2007) Psychometric characteristics of the Brazilian-Portuguese versions of the Functional Rating Index and the Roland Morris Disability Questionnaire. *Spine* 32:1902-1907.

Costa LOP, Maher CG, Latimer J, Ferreira PH, Ferreira ML, Pozzi GC, Freitas LMA (2008) Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil. Which one is the best? *Spine* 33:2459-2463.

Costa LOP, Maher CG, Latimer J, Hodges PW, Herbert RD, Refshauge KM, McAuley JH, Jennings MD (2009) Exercise for chronic low back pain: A randomised, placebo-controlled trial. Submitted for publication to *PLoS Medicine* on 16.03.09.

Costa LOP, Maher CG, Latimer J, Hodges PW, Shirley D (2009) An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain. Submitted for publication to *European Spine Journal* on 08.01.09.

Costa LOP, Maher CG, Latimer J, Smeets RJEM (2009) Reproducibility of real-time ultrasound for the measurement of abdominal muscle activity: a systematic review. Accepted for publication in *Physical Therapy* on 02.02.09.

Maher CG, Latimer J, Hodges PW, Refshauge KM, Moseley GL, Herbert RD, Costa LOP, McAuley J (2005) The effect of motor control exercise versus placebo in patients with chronic low back pain. *BMC Musculoskeletal Disorders* 6:54.

Maher CG, Latimer J, Costa LOP (2007) The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments. *Revista Brasileira de Fisioterapia* 11:245-252.

Presentations

Costa LOP, Maher CG, Latimer J (2006) Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations. *International Forum VIII, Primary Care Research on Low Back Pain*. Amsterdam, The Netherlands.

Costa LOP, Maher CG, Latimer J (2007) Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations. *15th International WCPT Congress*. Vancouver, Canada.

Costa LOP, Maher CG, Latimer J, Ferreira PH, Pozzi GC, Ribeiro RN (2007) Psychometric characteristics of the Brazilian-Portuguese versions of the Functional Rating Index and the Roland Morris Disability Questionnaire. *15th International WCPT Congress*. Vancouver, Canada.

Costa LOP, Maher CG, Latimer J, Hodges PW, Refshauge KM, Moseley GL, Herbert RD, McAuley J, Shirley D (2007) Lessons learnt from a randomised controlled trial of motor control exercise versus placebo in patients with low back pain. *Musculoskeletal Physiotherapy Australia 15th Biennial Conference*. Cairns, Australia.

Costa LOP, Maher CG, Latimer J, Hodges PW, Refshauge KM, Moseley GL, Herbert RD, McAuley J, Shirley D, Jennings MD (2007) The efficacy of motor control exercise versus placebo in patients with chronic low back pain: design issues and progress of a randomised controlled trial. *Research Showcase 2007*. Liverpool Hospital, Sydney, Australia.

Costa LOP, Maher CG, Latimer J, Smeets RJEM (2008) Reproducibility of real-time ultrasound for the measurement of abdominal muscle activity: a systematic review. *From cell to society 6, 6th Biennial Health Research Conference*. Leura, Australia.

Preface

This thesis is arranged in ten chapters, written so that each chapter is able to be read independently. The work in this thesis involves two topics: motor control exercise (Chapters 2-5) and clinimetrics (Chapters 6-9). The University of Sydney allows published papers that arose from the candidature to be included in the thesis.

Chapters 2, 6, 7, 8 and 9 are the PDF files of the published papers and Chapters 3, 4 and 5 are the submitted manuscripts.

Chapter One is an introduction to the thesis and provides an overview of the relevant low back pain literature. Chapter Two is the full published protocol for the randomised placebo-controlled trial described in Chapter Three. The paper is presented as published in *BMC Musculoskeletal Disorders*. Chapter Three describes the results of a randomised placebo-controlled trial investigating the efficacy of motor control exercise in patients with chronic low back pain. The paper is presented in the format required by *PLoS Medicine* where it was submitted for publication.

Chapter Four is a systematic review investigating the reproducibility of rehabilitative ultrasound imaging for the measurement of abdominal muscle activity. The paper is presented in the format required by *Physical Therapy* where it has been accepted for publication. Chapter Five is a reproducibility study of ultrasound measures of abdominal muscle activity in patients with chronic low back pain. The paper is presented in the format required by the *European Spine Journal* where it was submitted for publication.

Chapter Six is a narrative review about the relevance of cross-cultural adaptation and clinimetrics studies for self-report outcome measures. This paper is presented as published in *Revista Brasileira de Fisioterapia (Brazilian Journal of Physical Therapy)*.

Chapter Seven is a systematic review describing the available cross-cultural adaptations of low back pain self-report outcome measures and the clinimetric testing that has occurred for each adaptation. This paper is presented as published in *Spine*. Chapter Eight is a cross-cultural adaptation and clinimetric testing of the Brazilian-Portuguese versions of the *Functional Rating*

Index and the *Roland Morris Disability Questionnaire*. This paper is presented as published in *Spine*. Chapter Nine is a head-to-head comparison of the clinimetric properties of the Brazilian-Portuguese versions of the *Patient-Specific Functional Scale*, the *Functional Rating Index* and the *Roland Morris Disability Questionnaire*. This paper is presented as published in *Spine*. Finally, Chapter Ten consists of an overview, clinical implications and directions for further research.

Each chapter contains its own reference list. Appendices which were published as online supplementary material are included at the end of the relevant chapter. A copy of the “Guidelines for publication” for the journals *PLoS Medicine*, *Physical Therapy* and *European Spine Journal* are included at the end of the thesis. Ethical approval was gained from the Human Research Ethics Committee of the University of Sydney for all studies prior to commencement.

Abstract

Low back pain is a significant public health problem in many countries of the world being one of the major causes of work absence and disability. Although the outlook for evidence-based management of low back pain has greatly improved over the past decades, many questions remain. Questions related to treatment options, underlying mechanisms of treatment effects and optimal assessment of low back pain have yet to be fully addressed by researchers. The broad aim of this thesis therefore was to contribute to a better understanding of the contemporary management of low back pain by performing studies in these key research areas.

Most clinical practice guidelines recommend exercise as an effective treatment option for chronic low back pain. However the evidence for this recommendation comes from trials that are not placebo-controlled and so this may potentially provide biased estimates of the effects of exercise. Therefore a randomised controlled trial testing the effect of motor control exercise versus placebo in patients with chronic low back pain was conducted. Chapters 2 and 3 describe the trial protocol and the report of the trial respectively. A total of 154 patients with chronic low back pain were randomised to receive a motor control exercise program, or placebo (i.e. detuned short-wave therapy and detuned ultrasound therapy). Primary outcomes were pain, function, and the patient's global impression of recovery measured at 2 months. The exercise intervention improved function and patient's global impression of recovery, but not pain, at 2 months. The mean effect of exercise on function was 1.1 points (95%CI, 0.3 to 1.8), the mean effect on global impression of recovery was 1.5 points (95%CI, 0.4 to 2.5) and the mean effect on pain was 0.9 points (95%CI, -0.01 to 1.8), all measured on 11 point scales. Secondary outcomes also favoured

motor control exercise. This is the first study ever to demonstrate that motor control exercise is better than placebo for patients with chronic low back pain. Most of the treatment effects were maintained at 6 and 12 months follow-up. These results suggest that this intervention should be considered for patients with chronic low back pain in order to improve disability, function, and global impression of recovery, and to improve pain intensity in the long term, but not in the short term.

Rehabilitative ultrasound imaging (RUSI) has been increasingly used by physiotherapists in order to identify impairments in motor control as well as to monitor progress of patients with low back pain. As with any other clinical measure it is important to know how reproducible the RUSI measures are, and although there are some reproducibility studies in the literature, no systematic review on this topic has been conducted. Therefore a systematic review was performed with the objective of assessing the reproducibility studies of RUSI for abdominal wall muscles (Chapter 4). Eligible studies were identified via searches in CINAHL, EMBASE and MEDLINE with citation tracking via the Web of Science Index. A total of 21 studies were included. Due to heterogeneity of the studies' designs, pooling the data for a meta-analysis was not possible. RUSI measures of *thickness* of abdominal wall muscles were found to be reliable. Few studies analysed the reliability for the measurement of *thickness changes* (reflecting the muscle activity) finding good to poor results. Evidence for the reproducibility of the *difference in thickness changes over time* (necessary to evaluate improvements in muscle activity with treatment) was not available. A limitation of the existing literature is that studies typically had suboptimal designs and analysis. The current evidence for the reproducibility of RUSI for measuring abdominal muscle activity is mainly based upon studies with

suboptimal designs that included mostly healthy subjects, making generalisability to clinical settings uncertain.

Some questions about the reproducibility of RUSI measures of abdominal wall muscles are still unanswered; this is mainly due to design issues, such as inadequate statistics, inadequate sampling and lack of control of sources of bias (e.g. blinding and absence of controlling for ordering effects). In addition the clinically important questions about the reproducibility of *thickness changes* (reflecting the muscle activity) and *differences in thickness changes over time* (reflecting the improvement or deterioration of muscle activity) have not been adequately investigated. Therefore a reproducibility study that aimed to answer these questions was performed (Chapter 5). Thirty-five patients seeking care for chronic low back pain participated in this study. RUSI measures were taken at baseline and eight weeks post-baseline. Replicate measures of *thickness*, *thickness changes* and *differences in thickness changes over time* were analysed. The reproducibility of static images (*thickness*) was excellent ($ICC_{2,1} = 0.97$, 95% CI = 0.96-0.97, Standard Error of the Measurement (SEM) = 0.04cm, Smallest Detectable Change (SDC) = 0.11cm), the reproducibility of *thickness changes* was moderate ($ICC_{2,1} = 0.72$, 95% CI 0.65-0.76, SEM = 15%, SDC 41%), while the reproducibility of *differences in thickness changes over time* was poor ($ICC_{2,1} = 0.44$, 95% CI 0.33-0.58, SEM = 21%, SDC = 66.5%). Improvements in the test protocol should be undertaken in order to enhance the reproducibility of RUSI measures, especially for *differences in thickness changes over time*.

Self-report outcome measures (questionnaires) are widely used by health care providers for measuring patient's health status or treatment outcomes. Most of the questionnaires related to low back pain were developed in English and therefore their usefulness in non-English speaking countries is considerably limited. Cross-cultural adaptation and clinimetric testing are possibly the most efficient methods for solving this problem. Although there are many publications on the topic, a simple guide on how to perform a cross-cultural adaptation and clinimetric testing was not available. Therefore a "clinician-friendly" narrative review for Brazilian physical therapists (Chapter 6) was written. This review aimed firstly to explain the concepts and the relevance of cross-cultural adaptation and clinimetrics testing, secondly to summarise the current guidelines on the topic, thirdly to provide advice on how to choose a relevant questionnaire and finally how to evaluate the quality of an adapted questionnaire. Some examples of cross-cultural adaptations and clinimetrics testing of relevant low back pain questionnaires in the Brazilian-Portuguese language were also provided.

Although the number of international versions of low back questionnaires is growing, to date it is unclear which questionnaires have been cross-culturally adapted and into which specific language. To answer these questions a systematic review was conducted in order to describe the available cross-cultural adaptations of low back pain self-report outcome measures and the clinimetric testing that has occurred for each adaptation (Chapter 7). Searches were performed in MEDLINE, EMBASE, CINALH and LILACS; these searches were supplemented with information from experts in the field of low back pain from 27 different countries to ensure that the results were comprehensive. Sixty-one adaptations were identified.

While there are a large number of low back pain questionnaires available, very few have been adapted into other languages, particularly commonly spoken languages such as Mandarin, Hindi and Portuguese. The quality and comprehensiveness of clinimetric testing varied considerably, with the evaluation of reliability and construct validity most common. Further cross-cultural adaptation and clinimetric studies are clearly needed and special consideration must be given to study designs for clinimetric testing.

The final aim of this thesis was to cross-culturally adapt self-report instruments relevant to the management of low back pain in Brazil. This was achieved by two independent studies. The first study (Chapter 8) aimed to cross-culturally adapt the *Functional Rating Index* (FRI) into Brazilian-Portuguese and to test the clinimetric properties of the FRI and also of an existing Brazilian-Portuguese version of the *Roland Morris Disability Questionnaire* (RMDQ) which was not fully evaluated in the original study. Both instruments were tested for internal consistency, reliability, construct validity, ceiling and floor effects and internal responsiveness in 140 chronic low back patients presenting for physiotherapy treatment in Brazil. Both instruments were considered reliable and valid for the measurement of disability in Brazilian-Portuguese speakers with low back pain, no ceiling or floor effects were detected, but the internal responsiveness of both instruments was considered small.

The second study (Chapter 9) aimed to cross-culturally adapt the *Patient-Specific Functional Scale* (PSFS) and to perform a head-to-head comparison of the clinimetric properties of the PSFS, RMDQ and FRI. All instruments were tested for internal consistency, reliability, construct validity, ceiling and floor effects, internal

and external responsiveness in 99 acute low back patients presenting for physiotherapy treatment in Brazil. In order to fully test the construct validity and external responsiveness, it was necessary to cross-culturally adapt the *Pain Numerical Rating Scale* and the *Global Perceived Effect Scale*. The results of this study demonstrate that the Brazilian-Portuguese versions of the RMDQ, FRI and PSFS have similar clinimetric properties to each other and to the original English versions; however the PSFS was the most responsive instrument. The results from the studies in Chapters 8 and 9 will benefit the understanding of low back pain by enabling international comparisons between studies conducted in Brazil and English-speaking countries. In addition it will encourage researchers to include Brazilian-Portuguese speakers in their future clinical trials.

Overall, the studies included in this thesis have provided an important contribution to the contemporary management of low back pain. Firstly the use of motor control exercise could be considered for patients with chronic low back pain as it produces improvements in global impression of recovery, function, disability and pain. Secondly RUSI measures of abdominal wall muscles in patients with low back pain were considered reproducible for the measurement of muscle activity, but not as an outcome measure to detect improvement/deterioration of muscle activity over the course of treatment. Thirdly just a few high-quality cross-cultural adaptations and clinimetrics testing for self-report outcome measures relevant to the management of low back pain are available, and clearly more studies in this area are needed. Finally the Brazilian-Portuguese versions of the *Functional Rating Index*, the *Roland Morris Disability Questionnaire* and the *Patient-Specific Functional Scale* have acceptable

clinimetric properties and could be used in clinical practice as well as in research studies in Brazil.

Chapter One

Introduction

1.1 Introduction to low back pain

Low back pain is a major health and socioeconomic problem and is associated with high costs in care, work absenteeism and disability worldwide^{1,2}. A systematic review of prevalence studies reported estimates for point prevalence of low back pain ranging from 12% to 33%, the one-year prevalence ranging from 22% to 65% and the lifetime prevalence ranging from 11% to 84%³. Small differences were found for these estimates in patients from African countries⁴. This wide range of estimates could be explained by differences in study designs and differences in low back pain definitions³.

1.2 Economic burden of low back pain

In the context of such a prevalent problem, low back pain is now considered a major public health problem in many countries such as the United States⁵, Australia⁶ and European countries¹. The costs associated with low back pain are substantial, with the costs in Australia being approximately A\$1 billion a year⁶, while in the United States the total annual cost associated with low back pain has reached 50 billion^{5,7}. In Europe approximately 90% of the total costs are indirect costs due to work absenteeism and disablement. Although the economic costs vary among different countries⁸, it is clear that low back pain represents an important economic burden worldwide.

1.3 Low back pain classification

Low back pain is defined as pain and discomfort, localised below the costal margin and above the inferior gluteal folds, with or without referred leg pain¹. Low back pain is often classified using a diagnostic triage approach where patients enter one of

three categories: non-specific low back pain, serious spinal pathology and nerve root compromise⁹. Nerve root compromise can arise from a disc prolapse, spinal stenosis or surgical scarring and can be identified by a careful neurological examination. Serious spinal pathology includes diseases such as tumours, fractures, infection and inflammatory diseases such as ankylosing spondylitis. Finally the remaining patients are classified as having non-specific low back pain. Most of the low back pain cases are classified as non-specific, less than 1% are due to serious spinal pathology and less than 5% are true nerve root compromise⁹. For the rest of this thesis the term *non-specific low back pain* will be used interchangeably with *low back pain*.

Low back pain is often classified in three stages (acute, sub-acute and chronic) according to its duration and this provides some information to the clinician with regards to treatment and prognosis. Acute low back pain is usually defined as an episode persisting for less than 6 weeks; sub-acute low back pain as low back pain persisting between 6 to 12 weeks and chronic low back pain as low back pain persisting for 12 weeks or longer¹.

Another approach to classification of low back pain is the use of treatment-based classification such as those advocated by McKenzie¹⁰ and by Fritz and colleagues^{11,12}. The McKenzie method was based on Robin McKenzie's clinical observations of patients with low back pain¹⁰. The McKenzie classification assessment involves history-taking and a screening examination for serious spinal pathology followed by a comprehensive clinical examination consisting of an analysis of posture, range of movement and an assessment of symptom response to different loading strategies. The combination of the history-taking and physical

assessment allows the therapist to classify the patient into one of three syndromes (derangement, dysfunction and postural)¹⁰. The treatment is then tailored for each specific syndrome. Another low back pain classification system was developed by Dellito and colleagues¹³ which was updated later by Fritz and colleagues^{11,12}. Similarly to the McKenzie method, this classification system involves screening patients for serious spinal pathologies followed by history-taking and physical examination to place a patient into one of four treatment-based categories: manipulation, specific exercise (flexion, extension, and lateral shift patterns), stabilisation and traction.

1.4 The prognosis of low back pain

Clinical guidelines suggest that acute low back pain has an extremely favourable prognosis, for example the UK guideline states that 90% of patients will recover completely in 6 weeks¹⁴. A more accurate (but less optimistic) picture about the prognosis of acute low back pain is now available from a recent systematic review¹⁵. This review concluded that pain decreased rapidly within one month (pooled mean 58%) with pain intensity levels continuing to decrease slowly up to three months and then remained constant until 12 months. Additionally it was observed that most people will have at least one recurrence within 12 months¹⁵. A similar figure was observed in a large inception cohort study of patients with acute low back pain¹⁶ in which 39% of the patients recovered completely in 6 weeks. By 12 weeks the probability of recovery was 57.4% increasing to 71.8% by 12 months. This study also identified that older age, back pain associated with compensation cases, higher pain intensity, longer duration of low back pain before consultation, more days of reduced activity because of low back pain before consultation, feelings of depression

and a perceived risk of pain becoming persistent were associated with poorer prognosis¹⁶.

The prognosis of chronic low back pain is, to date, poorly investigated. Most of the available studies about prognosis of chronic low back pain have enrolled unrepresentative samples that were not recruited at the onset of the low back pain episode^{17,18} or have large losses to follow up^{19,20}. A recent inception cohort study of 406 chronic low back pain patients found that thirty-five percent of patients had recovered completely within 9 months and forty-one percent of patients recovered from their initial episode of chronic low back by 12 months. Despite the common view that recovery from an episode of chronic low back pain is unlikely, this study shows that an important proportion of patients recovered within a period of one year. Patients with higher disability levels at the onset of chronicity, previous sick leave due to low back pain, lower levels of education and higher perceived risk of persistent pain were more likely to have delayed recovery from their pain. Those born overseas (i.e. born outside Australia), with high disability levels at the onset of chronicity and higher perceived risk of persistent pain were more likely to have delayed recovery in terms of disability²¹.

1.5 Treatment possibilities for low back pain

There is a wide range of treatment possibilities for patients with low back pain, including education²², behavioural therapy²³, medication²⁴⁻²⁷, physical modalities²⁸⁻³¹, manual therapy³², exercise³³ and others. One factor that may be considered when choosing the best treatment is the duration of low back pain¹. There is a consensus from guidelines that patients with acute low back pain should receive advice about

the favorable prognosis of this condition and to remain active as much as possible. Additionally education booklets and pain management are also indicated^{1,2}. For subacute low back pain, guidelines recommend interdisciplinary rehabilitation and functional restoration as evidence-based options². A recently published randomised placebo-controlled trial demonstrated that a combination of supervised exercises and advice was also efficacious for subacute low back pain³⁴. Finally for chronic low back pain, the European Guidelines¹ recommend a range of conservative options such as cognitive behavioural therapy, exercise, educational interventions and multidisciplinary treatment.

Another approach to choosing the best treatment for an individual patient is to classify the patients into groups and treat them accordingly. This approach has been suggested by some authors such as McKenzie¹⁰ and Fritz¹². Although there is some preliminary evidence available about the effectiveness of the McKenzie classification system³⁵, a systematic review concluded that the effectiveness of this method is yet to be established³⁶. Fritz and colleagues have been testing and adjusting another classification method^{11,12} which involves classification of patients with low back pain in 4 different categories (traction, stabilisation, specific exercise and manipulation). The results from a range of studies using this treatment classification approach provide some indication that assigning patients to subgroups and treating them accordingly could provide better results than a generic approach to treatment³⁷⁻⁴².

1.5.1 Exercise as a treatment for chronic low back pain

Exercise is possibly the most advocated treatment option for chronic low back pain^{1,2,43}. This is not a surprise given the fact that there is a good biological rationale

for the use of exercise for patients with chronic low back pain (i.e. exercise has the potential to improve spinal mobility, muscle strength, motor coordination, spinal stabilisation and general aerobic conditioning⁹). Additionally, the costs for exercise therapy are reasonably low and exercise providers are readily available.

The latest Cochrane Review of exercise for the treatment of low back pain³³ identified 43 randomised controlled trials specifically dedicated to the management of chronic low back pain. This review found strong evidence that exercise therapy is at least as effective as other interventions and conflicting evidence that exercise therapy is more effective than other treatments for chronic low back pain. The effects on pain and functioning were considered clinically meaningful in studies from healthcare populations (i.e. patients seeking care for low back pain) in which improvements were significantly greater than those observed in studies from general or mixed populations (i.e. patients recruited from the community, advertisements, etc).

The conclusions from randomised controlled trials and systematic reviews of exercise for low back pain (such as the Cochrane review³³) have influenced most of the available clinical practice guidelines that consistently recommend exercise as an important component of the treatment of chronic low back pain^{1,2,14,43}. Although this sounds promising it is important to state that the recommendations from the guidelines are based heavily on trials that compare exercise against treatments with unknown efficacy⁴⁴. Moreover limitations in the quality and reporting of trials are notable³³. These features can overestimate the effects of exercise and therefore high-quality trials are still needed. This need is confirmed by the European Guideline¹

which recommends that “*effectiveness of specific types of exercise therapy needs to be further evaluated. This includes the evaluation of spinal stabilisation exercises, McKenzie exercises and other popular exercise regimens that are often used but inadequately researched (p.S196)*”.

1.5.2 Motor control exercise for chronic low back pain

Motor control exercise (also known as specific stabilisation exercise) was first considered as a treatment for low back pain about 13 years ago, when a group of researchers from The University of Queensland in Australia published the first manuscript on this topic⁴⁵. Since then the number of studies about this topic has increased⁴⁶⁻⁴⁸ as well as its popularity and use in clinical practice.

The biological rationale of motor control exercise is fundamentally based on the idea that the stability and control of the spine are altered in people with low back pain⁴⁵. Physiological studies have demonstrated that patients with low back pain may present with a delayed onset of activity of the deep trunk muscles (such as the Transversus abdominis and Multifidus) when the stability of the spine is challenged in dynamic tasks^{49,50}. Morphologically a lower cross sectional area⁵¹ and larger percentage of intramuscular fat in the multifidus muscle⁵² were found in patients with low back pain compared to asymptomatic controls. Moreover it was found that patients with low back pain tend to increase the spinal stiffness to compensate for the lack of stability from the deep muscles by increasing the activity of the superficial muscles⁵³. Finally it was demonstrated that patients who recovered from an episode of acute low back pain are more susceptible to recurrence and chronicity if the above changes were not treated with motor control exercise⁵⁴.

Given the findings from physiological studies, a treatment model was developed⁵⁵. Firstly patients should be comprehensively assessed for suitability. If suitable the clinician would prescribe an individualised exercise program based on the patient's presentation. The first stage of the treatment would be to retrain the muscles whose role is considered to be that of a primary spine stabiliser (e.g. Transversus abdominis and Multifidus)⁵⁵⁻⁵⁷. These exercises are usually supplemented with exercises for the pelvic floor muscles, breathing control and control of spinal posture and movement^{55,58}. Patients should be taught to contract the deep muscles independently without substitution of the superficial muscles. The progression of the training is made by starting from easier positions (e.g. supine or four-point kneeling) progressing to harder ones such as sitting and standing. Physiotherapists should be able to stop the exercises as soon as any substitution from the superficial muscles occurs, breathing control is lost, muscle fatigue occurs or pain intensity increases⁵⁶. It has been suggested that when the patient is able to sustain an isolated contraction of the deep muscles 10 times for at least 10 seconds each, a progression for the second stage should be made⁵⁵.

The second stage involves most of the principles of the first stage, but a more dynamic approach is now emphasized⁵⁵. Functional activities that stimulate trunk coordination in association with limb movement (such as walking, reaching and work-related movements) are encouraged^{55,56}. Patients must be supervised on an individual basis in order to monitor the exercises properly as well as the progression of the treatment. Home exercises are prescribed during the whole course of treatment and ideally patients should be able to perform the functional exercises while contracting the muscles that control and support the spine^{55,56}.

1.5.3 Efficacy of motor control exercise for chronic low back pain

It seems logical that the presence of a number of laboratory studies evaluating mechanisms of action of motor control exercise and its popularity in clinical practice would stimulate studies evaluating the effectiveness of this type of intervention. In fact a good number of clinical trials on this topic have been performed and three systematic reviews are now available⁴⁶⁻⁴⁸. The most recent systematic review is confined to clinical trials of motor control exercise for chronic low back pain patients⁴⁷ and as an advantage from the two previous systematic reviews^{46,48} a meta-analysis approach was used. This review identified thirteen randomised controlled trials and one quasi-randomised controlled trial, all being studies comparing motor control exercise against other treatments (such as spinal manipulative therapy, other exercise regimes, education, surgery, etc) or no treatment. Notably no placebo-controlled trials were identified. The conclusions from this review were that motor control exercise is superior to minimal intervention and benefits of adding motor control exercise to another therapy were observed. However motor control exercise is not more effective than manual therapy or other forms of exercise in patients with chronic low back pain⁴⁷.

To date there are no trials of exercise for chronic low back pain patients that are placebo-controlled^{33,47}. This type of study design has the advantage of providing the least biased estimates of effects of the intervention⁵⁹ as it controls for important sources of bias such as placebo effects, change of patient behavior caused by knowledge of allocation, measurement bias, treatment non-compliance and loss to follow up⁶⁰. As a consequence, a high-quality randomised placebo-controlled trial of exercise for patients with chronic low back pain is needed. Chapter 2 presents the

published study protocol of a randomised placebo-controlled trial of motor control exercise in patients with chronic low back pain and the results of this trial are described in Chapter 3.

1.6 The assessment of motor control changes

Many of the treatment outcomes in clinical practice are measured by self-report instruments. Some features, like flexibility and strength, are usually measured by instruments such as goniometers and dynamometers. The reference standard for measuring motor control changes is fine-wire electromyography, which has been used in laboratory studies to identify motor control patterns in patients with low back pain and asymptomatic controls^{45,49,61,62}. However, as an invasive procedure, electromyography is painful, uncomfortable and has associated risks such as infection. Additionally electromyography is costly. All these features make its use in clinical practice and in clinical studies difficult.

1.6.1 Rehabilitative Ultrasound Imaging

Rehabilitative Ultrasound Imaging (RUSI)⁶³ has been used as a non-invasive alternative to electromyography as abdominal muscle activity could be indirectly measured by morphologic changes (thickness changes)⁶⁴. There is evidence that RUSI measures of thickness changes are correlated with electromyography measures of muscle activity at low levels of contraction force⁶⁵ and also that RUSI measures are able to distinguish patients with low back pain from normal controls⁶⁶. The results of these studies provide some indication of the validity of RUSI as an indirect measurement of abdominal wall muscle activity.

1.6.2 Reproducibility of Rehabilitative Ultrasound Imaging

Reproducibility can be defined as the degree to which repeated measurements in stable persons (in a test-retest design) provide similar answers⁶⁷. Reproducibility is also considered an umbrella term for reliability⁶⁸ (i.e. the extent to which patients can be distinguished from each other, despite measurement errors - relative measurement error) and agreement⁶⁸ (i.e. the extent to which the scores on repeated measures are close to each other - absolute measurement error). There are different ways for considering reproducibility of RUSI measures. The first and most common, is to measure the reproducibility of single muscle thickness. The second is to measure muscle activity by *thickness changes* and finally to measure the improvement or deterioration of muscle activity by *differences in thickness changes*. Several studies aimed to determine the reproducibility of RUSI measures of abdominal wall muscles have been conducted; however there is no systematic review available on this topic. Chapter 4 describes a systematic review of reproducibility studies of RUSI measures of abdominal wall muscles.

The available studies of RUSI measures answer some, but not all, of the questions about its reproducibility. Most of the studies have recruited young and healthy participants and just three studies recruited patients with low back pain⁶⁹⁻⁷¹, in addition numerous limitations in terms of design and statistical analysis were observed⁷². Finally the systematic review in Chapter 4 revealed that there is no study in the literature that aimed to investigate the reproducibility of *differences in thickness changes* over longer periods of time. Therefore a well designed reproducibility study of RUSI measures of abdominal muscle activity in a clinical

relevant population of low back pain patients was needed. This study is described in Chapter 5 of this thesis.

1.7 Self-report outcome measures for low back pain

Clinicians and researchers commonly use questionnaires to assess treatment outcomes⁷³, to estimate prognosis⁷⁴, to collect information about patient's feelings or thoughts⁷⁵ and for screening purposes (for example screening for contra-indications for exercise prescription⁷⁶). Questionnaires are convenient, cheap and in most cases their measurement (clinimetric) properties are superior to measures made in the physical examination⁷⁷. In 1997 a panel of experts made a series of recommendations about the use and the standardisation of self-report outcome measures for low back pain⁷⁸. Since then the use of this type of measurement has increased considerably in clinical research.

1.7.1 Availability of self-report measures for low back pain

Most of the available self-report outcome measures for low back pain were developed in English-speaking countries. In fact just four (out of forty) low back pain questionnaires were developed in a language other than English⁷⁹. This is not a surprise as English-speaking countries contribute substantially to research output; on the other hand just 6-7% of the whole population speaks English as their first language. This dilemma raised some questions such as:

1. "How clinicians and researchers from non-English speaking countries would be able to assess their patients or research participants properly?"

2. “How can the results of studies from non-English speaking countries be compared against studies from English-speaking countries?”
3. “How can we avoid the exclusion of non-English speakers from clinical trials performed in countries like the United States, Canada or Australia, where a substantial number of residents speak languages other than English?”

Possibly the most efficient way to solve these issues is to choose an existing popular and relevant questionnaire (for example, the *Roland Morris Disability Questionnaire*⁷³ or the *Oswestry Disability Index*⁸⁰) and then cross-culturally adapt it into the target language and culture. A consensus about how to perform studies on cross-cultural adaptation is easily available from two guidelines on this topic^{81,82}. But this is only the first step; the second step (and certainly the most complicated) is to check if the new version retains similar clinimetric properties (such as reliability, validity and responsiveness) to the original questionnaire. Chapter 6 presents a narrative review on the relevance of cross-cultural adaptation studies and clinimetric testing, it also offers examples of international versions relevant to low back pain that were adapted and tested as suggested by the current guidelines of cross-cultural adaptation^{81,82} and clinimetric testing⁶⁷.

A recent systematic review of back-specific outcome questionnaires⁸³ has identified thirty six different instruments dedicated to monitoring outcomes in patients with low back pain. Most of these questionnaires were classified as having acceptable clinimetric properties. However it is not clear which of the low back pain-related questionnaires were cross-culturally adapted and in which languages. Chapter 7 presents a systematic review that aimed to describe the cross-cultural adaptations

relevant to the management of low back pain that are available and to describe the clinimetric testing that occurred for each adaptation.

1.7.2. Cross-cultural adaptation and clinimetric studies in Brazil

One of the conclusions of Chapter 7 was that there was no association between the number of people who speak a specific language and the number of adaptations to that language⁷⁹. One interesting example is Brazil, a country with a population of more than 180 million who speak Portuguese as their first language. Portuguese is the sixth most spoken language in the world however there is just one low back pain related questionnaire (the *Roland Morris Disability Questionnaire*⁸⁴) adapted in Portuguese. Brazil is currently in 17th place in the world ranking of scientific publications⁸⁵ with a great potential for developing studies in low back pain and it could be argued that this research gap must be filled. Chapter 8 presents a cross-cultural adaptation and clinimetric testing of two self-report outcome measures for low back pain: *The Roland Morris Disability Questionnaire*⁷³ and the *Functional Rating Index*⁸⁶. A second cross-cultural adaptation study with a head-to-head comparison of clinimetric properties of three self report outcome measures for low back pain in Brazil was also undertaken and is described in Chapter 9.

1.8 Aims of the thesis

The broad aim of this thesis was to contribute to a better understanding of the contemporary management of low back pain through a series of high-quality studies. The specific aims of this thesis were to:

1. Conduct a randomised controlled trial to confirm whether motor control exercise is better than placebo for the treatment of patients with chronic low back pain (Chapters 2 and 3).
2. Perform a systematic review to summarise the current evidence of the reproducibility of Rehabilitative Ultrasound Imaging for the measurement of abdominal wall muscle activity (Chapter 4).
3. Determine the reproducibility for static images, muscle activity and improvement/deterioration of muscle activity measured by Rehabilitative Ultrasound Imaging in patients seeking care for their chronic low back pain (Chapter 5).
4. Produce a “clinician-friendly” review about the relevance of cross-cultural adaptation and clinimetric testing of self-report outcome measures (Chapter 6).
5. Identify the available cross-cultural adaptations of self-report outcome measures for low back pain and describe the clinimetric testing that has occurred for each adaptation (Chapter 7).
6. Cross-culturally adapt the following low back pain-related questionnaires into Brazilian-Portuguese: *Functional Rating Index*, *Patient-Specific Functional Scale*, *Global Perceived Effect* and *Pain Numerical Rating Scale* and test the clinimetric properties of the *Functional Rating Index*, *Patient-Specific Functional Scale*, *Global Perceived Effect*, *Pain Numerical Rating Scale* and the *Roland Morris Disability Questionnaire* in patients with low back pain (Chapters 8 and 9).

1.9 References

1. Cost B13 working group. European Guidelines for the Management of Chronic Non-Specific Low back Pain. *Eur Spine J* 2006;15 S192-S300.
2. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: A joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478-91.
3. Walker BF. The prevalence of low back pain: A systematic review of the literature from 1966 to 1998. *J Spinal Disord* 2000;13:205-17.
4. Louw QA, Morris LD, Grimmer-Somers K. The prevalence of low back pain in Africa: a systematic review. *BMC Musculoskelet Disord* 2007;8:105.
5. Deyo RA. Low back pain. *Sci Am* 1998;279:48-53.
6. Walker B, Muller R, Grant W. Low back pain in Australian adults: the economic burden. *Asia Pac J Public Health* 2003;15:79-87.
7. Pai S, Sundaram LJ. Low back pain: an economic assessment in the United States. *Orthop Clin N Am* 2004;35:1-5.
8. Dagenais DC, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *The Spine J* 2008;8:8-20.
9. Waddell G. *The Back Pain Revolution*. Churchill Livingstone, 2004.
10. McKenzie RA, May S. *The lumbar spine: mechanical diagnosis and therapy*. Waikanae, New Zealand: Spinal Publications, 2003.
11. Fritz JM, Brennan GP, Clifford SN, et al. An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. *Spine* 2006;31:77-82.
12. Fritz JM, Cleland JA, Childs JD. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *J Orthop Sports Phys Ther* 2007;37:290-302.
13. Delitto A, Erhard RE, Bowling RW, et al. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther* 1993;75:470-85.

14. Koes BW, Van Tulder MW, Ostelo RWJG, et al. Clinical guidelines for the management of low back pain in primary care: an international comparison. *Spine* 2001;26:2504-14.
15. Pengel LHM, Herbert RD, Maher CG, et al. Acute low back pain: systematic review of its prognosis. *BMJ* 2003;327:323.
16. Henschke N, Maher CG, Refshauge KM, et al. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ* 2008;337:a171.
17. van Tulder MW, Koes BW, Metsemakers JF, et al. Chronic low back pain in primary care: a prospective study on the management and course. *Fam Pract* 1998;15:126-32.
18. van Tulder MWP, Koes BWP, Bouter LMP, et al. Management of Chronic Nonspecific Low Back Pain in Primary Care: A Descriptive Study. *Spine* 1997;22:76-82.
19. Grotle M, Vollestad NK, Brox JI. Clinical course and impact of fear-avoidance beliefs in low back pain. Prospective cohort study of acute and chronic low back pain: II. *Spine* 2006;31:1038-46.
20. Carey TS, Garrett JM, Jackman AM. Beyond the good prognosis. Examination of an inception cohort of patients with chronic low back pain. *Spine* 2000;25:115-20.
21. Costa LCM, Maher CM, McAuley JM, et al. Prognosis in patients with chronic low back pain: inception cohort study *Submitted to Lancet* 2009.
22. Engers A, Jellema P, Wensing M, et al. Individual patient education for low back pain. *Cochrane Database Syst Rev* 2008.
23. Ostelo R, Van Tulder MW, Vlaeyen JWS, et al. Behavioural treatment for chronic low-back pain. *Cochrane Database Syst Rev* 2005.
24. Deshpande A, Furlan A, Mailis-Gagnon A, et al. Opioids for chronic low-back pain. *Cochrane Database Syst Rev* 2007.
25. Gagnier JJ, vanTulder M, Berman B, et al. Herbal medicine for low back pain. *Cochrane Database Syst Rev* 2006.
26. Roclofs P, Deyo RA. Non-steroidal anti-inflammatory drugs for low back pain. *Cochrane Database Syst Rev* 2008.

27. Staal JB, de Bie R, de Vet HCW, et al. Injection therapy for subacute and chronic low-back pain. *Cochrane Database Syst Rev* 2008.
28. Clarke JA, Van TMW, Blomberg SEI, et al. Traction for low-back pain with or without sciatica. *Cochrane Database Syst Rev* 2007.
29. French SD, Cameron M, Walker BF, et al. Superficial heat or cold for low back pain. *Cochrane Database Syst Rev* 2006.
30. Khadilkar A, Odebiyi DO, Brosseau L, et al. Transcutaneous electrical nerve stimulation (TENS) versus placebo for chronic low-back pain. *Cochrane Database Syst Rev* 2008.
31. Ypusefi-Nooraie R, Schonstein E, Heidari K, et al. Low level laser therapy for nonspecific low-back pain. *Cochrane Database Syst Rev* 2008.
32. Assendelft WJJ, Morton SC, Yu EI, et al. Spinal manipulative therapy for low-back pain. *Cochrane Database Syst Rev* 2004.
33. Hayden JA, van Tulder MV, Malmivaara A, et al. Exercise therapy for treatment of non-specific low back pain. *Cochrane Database Syst Rev* 2005.
34. Pengel LHM, Refshauge KM, Maher CG, et al. Physiotherapist-directed exercise, advice or both for subacute low back pain: a randomized trial. *Ann Intern Med* 2007;146:787-96.
35. Long A, Donelson R, Fung T. Does it matter which exercise? A randomized controlled trial of exercise for low back pain. *Spine* 2004;29:2593-602.
36. Machado LAC, De Souza MVS, Ferreira PH, et al. The McKenzie method for low back pain. A systematic review of the literature with a meta-analysis approach. *Spine* 2006;31:E254-E62.
37. Browder DA, Childs JD, Cleland JA, et al. Effectiveness of an extension-oriented treatment approach in a subgroup of subjects with low back pain: A randomized clinical trial. *Phys Ther* 2007;87:1608-18.
38. Fritz JM, Lindsay W, Matheson JW, et al. Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. *Spine* 2007;32:E793-E800.
39. Brennan GP, Fritz JM, Hunter SJ, et al. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain - Results of a randomized clinical trial. *Spine* 2006;31:623-31.

40. Cleland JA, Fritz JM, Whitman JM, et al. The use of a lumbar spine manipulation technique by physical therapists in patients who satisfy a clinical prediction rule: A case series. *J Orthop Sports Phys Ther* 2006;36:209-14.
41. Cleland JA, Fritz JM, Childs JD, et al. Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who satisfy a clinical prediction rule: Study protocol of a randomized clinical trial [NCT00257998]. *BMC Musculoskelet Disord* 2006;7:11.
42. Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: A validation study. *Ann Intern Med* 2004;141:920-8.
43. Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: A review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:492-504.
44. Galantini ML, Bzedewka TM, Eissler-Russo JL, et al. The impact of modified Hatha yoga on chronic low back pain: a pilot study. *Altern Ther Health Med* 2004;10:56-9.
45. Hodges PW, Richardson CA. Inefficient muscular stabilization of the lumbar spine associated with low back pain. A motor control evaluation of transversus abdominis. *Spine* 1996;21:2640-50.
46. Ferreira PH, Ferreira ML, Maher CG, et al. Specific stabilisation exercise for spinal and pelvic pain: A systematic review. *Aust J Physiother* 2006;52:79-88.
47. Macedo LG, Maher CG, Latimer J, et al. Motor control exercises for persistent nonspecific low back pain: A systematic review. *Phys Ther* 2009;89:9-25.
48. Rackwitz B, de Bie R, Limm H, et al. Segmental stabilizing exercises and low back pain. What is the evidence? A systematic review of randomized controlled trials. *Clin Rehabil* 2006;20:553-67.
49. Hodges PW, Richardson CA. Delayed postural contraction of transversus abdominis in low back pain associated with movement of the lower limb. *J Spinal Disord* 1998;11:46-56.
50. Hodges PW, Richardson CA. Altered trunk muscle recruitment in people with low back pain with upper limb movement at different speeds. *Arch Phys Med Rehabil* 1999;80:1005-12.
51. Hides JA, Stokes MJ, Saide M, et al. Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain. *Spine* 1994;19:165-72.

52. Alaranta H, Tallroth K, Soukka A, et al. Fat-content of lumbar extensor muscles and low-back disability - a radiographic and clinical comparison. *J Spinal Disord* 1993;6:137-40.
53. van Dieen JH, Selen LP, Cholowicki J. Trunk muscle activation in low back pain patients, an analysis of literature. *J Electromyogr Kinesiol* 2003;13:333-51.
54. Hides JA, Jull GA, Richardson CA. Long-term effects of specific stabilizing exercises for first-episode low back pain. *Spine* 2001;26:E243-8.
55. Richardson CA, Jull GA, Hodges PW, et al. *Therapeutic Exercise for Spinal Segmental Stabilization in Low Back Pain*. Edinburg: Churchill Livingstone, 1999.
56. O'Sullivan PB. Lumbar segmental 'instability': clinical presentation and specific stabilizing exercise management. *Man Ther* 2000;5:2-12.
57. Twomey L, Taylor J. *Physical therapy for low back pain. Clinics in Physical Therapy*. New York: Churchill Livingstone, 1994.
58. Hodges PW, Ferreira PH, Ferreira ML. Lumbar spine: Treatment of instability and disorders of movement control. In Magee DJ, Zachazewski JE, Quillen WS eds. *Scientific Foundations and Principles of Practice in Musculoskeletal Rehabilitation*: Saunders, 2007.
59. Machado LAC, Kamper S, Herbert RD, et al. Analgesic effects of treatments for non-specific low back pain: a meta-analysis of placebo-controlled randomized trials. *Rheumatology* 2009;in press.
60. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Inter Med* 2002;136:254-9.
61. Hodges PW, Moseley GL. Pain and motor control of the lumbopelvic region: effect and possible mechanisms. *J Electromyogr Kinesiol* 2003;13:361-70.
62. Hodges PW, Richardson CA. Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Exp Brain Res* 1997;114:362-70.
63. Teyhen DS, Gill NW, Whittaker JL, et al. Rehabilitative ultrasound imaging of the abdominal muscles. *J Orthop Sports Phys Ther* 2007;37:450-66.
64. Hodges PW, Pengel LH, Herbert RD, et al. Measurement of muscle contraction with ultrasound imaging. *Muscle Nerve* 2003;27:682-92.

65. McMeeken JM, Beith ID, Newham DJ, et al. The relationship between EMG and change in thickness of transversus abdominis. *Clin Biomech* 2004;19:337-42.
66. Ferreira PH, Ferreira ML, Hodges PW. Changes in recruitment of the abdominal muscles in people with low back pain - Ultrasound measurement of muscle activity. *Spine* 2004;29:2560-6.
67. Terwee CB, Bot SDM, Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
68. de Vet HCW, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033-9.
69. Ferreira PH, Ferreira ML, Maher CG, et al. Clinical ultrasound test for transversus abdominis thickness: investigation of reliability 13th Biennial Conference - Musculoskeletal Physiotherapy Australia. Sydney, Australia, 2003.
70. Mannion AF, Pulkovski N, Gubler D, et al. Muscle thickness changes during abdominal hollowing: an assessment of between-day measurement error in controls and patients with chronic low back pain. *Eur Spine J* 2008;17:494-501.
71. Teyhen DS, Miltenberger CE, Deiters HM, et al. The use of ultrasound imaging of the abdominal drawing-in maneuver in subjects with low back pain. *J Orthop Sports Phys Ther* 2005;35:346-55.
72. Costa LOP, Maher CG, Latimer J, et al. An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain. *Eur Spine J* 2009;under review.
73. Roland M, Morris R. A study of natural history of back pain, part 1: development of a reliable and sensitive measure of disability in low back pain. *Spine* 1983;8:141-4. *Spine* 1983;8:145-50.
74. Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain* 1998;14:209-15.
75. Waddell G, Newton M, Henderson I, et al. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157-68.
76. ACSM. *ACSM's Guidelines for exercise testing and prescription*. Baltimore: Williams and Wilkins, 1995.

77. Pengel LHM, Refshauge KM, Maher CG. Responsiveness of pain, disability and physical impairment outcomes in patients with low back pain. *Spine* 2004;29:879-83.
78. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine*. 1998;23:2003-13.
79. Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations. *Spine* 2007;32:1028-37.
80. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine* 2000;25:2940-53.
81. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186-91.
82. Guillemin F, Bombardier C, Beaton DE. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417-32.
83. Grotle M, Brox JI, Vøllestad NK. Functional Status and Disability Questionnaires: What Do They Assess? A Systematic Review of Back-Specific Outcome Questionnaires. *Spine* 2004;30:130-40.
84. Nusbaum L, Natour J, Ferraz MB, et al. Translation, adaptation and validation of the Roland Morris questionnaire - Brazil Roland Morris. *Braz J Med Biol Res* 2001;34:203-10.
85. Thompson ISI. Top 20 country rankings in all essential science indicators, 2007. Available from: <http://in-cites.com/countries/2007allfields.html>. Accessed 12/02/2009.
86. Feise RJ, Menke MJ. Functional Rating Index - A new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine* 2001;26:78-87.

Chapter Two

The effect of motor control exercise versus placebo in patients with chronic low back pain

Chapter Two is published as:


Maher CG, Latimer J, Hodges PW, Refshauge KM, Moseley GL, Herbert RD, Costa LOP, McAuley J (2005) The effect of motor control exercise versus placebo in patients with chronic low back pain. *BMC Musculoskeletal Disorders* 6:54.

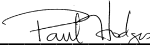
**Statement from co-authors confirming authorship contribution of the
PhD candidate**

As co-authors of the paper “The effect of motor control exercise versus placebo in patients with chronic low back pain”, we confirm that Leonardo Costa has made the following contributions:


- Conception and design of the research
- Writing of the manuscript and critical appraisal of the content


Christopher G Maher _____  _____ Date 20.03.2009

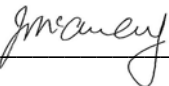
Jane Latimer _____  _____ Date 20.03.2009

Paul W Hodges _____  _____ Date 20.03.2009

Kathryn Refshauge _____ Date 20.03.2009

G Lorimer Moseley _____  _____ Date 20.03.2009

Robert D Herbert _____  _____ Date 20.03.2009

James McAuley _____  _____ Date 20.03.2009

Study protocol

Open Access

The effect of motor control exercise versus placebo in patients with chronic low back pain [ACTRN012605000262606]

Chris G Maher^{†1}, Jane Latimer^{†1}, Paul W Hodges^{†2}, Kathryn M Refshauge^{†1}, G Lorimer Moseley^{†3}, Robert D Herbert^{†1}, Leonardo OP Costa^{*†1} and James McAuley^{†1}

Address: ¹Back Pain Research Group, School of Physiotherapy, The University of Sydney, PO Box 170, Lidcombe, NSW, 1825, Australia, ²Division of Physiotherapy, The University of Queensland, Brisbane Qld 4072, Australia and ³Department of Human Anatomy & Genetics, Oxford University, South Parks Rd, Oxford, OX1 3QX, UK

Email: Chris G Maher - c.maher@fhs.usyd.edu.au; Jane Latimer - j.latimer@fhs.usyd.edu.au; Paul W Hodges - p.hodges@uq.edu.au; Kathryn M Refshauge - k.refshauge@fhs.usyd.edu.au; G Lorimer Moseley - lorimer.moseley@anat.ox.ac.uk; Robert D Herbert - r.herbert@fhs.usyd.edu.au; Leonardo OP Costa* - lcos3060@usyd.edu.au; James McAuley - j.mcauley@fhs.usyd.edu.au

* Corresponding author †Equal contributors

Published: 04 November 2005

Received: 29 September 2005

BMC Musculoskeletal Disorders 2005, 6:54 doi:10.1186/1471-2474-6-54

Accepted: 04 November 2005

This article is available from: <http://www.biomedcentral.com/1471-2474/6/54>

© 2005 Maher et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: While one in ten Australians suffer from chronic low back pain this condition remains extremely difficult to treat. Many contemporary treatments are of unknown value. One potentially useful therapy is the use of motor control exercise. This therapy has a biologically plausible effect, is readily available in primary care and it is of modest cost. However, to date, the efficacy of motor control exercise has not been established.

Methods: This paper describes the protocol for a clinical trial comparing the effects of motor control exercise versus placebo in the treatment of chronic non-specific low back pain. One hundred and fifty-four participants will be randomly allocated to receive an 8-week program of motor control exercise or placebo (detuned short wave and detuned ultrasound). Measures of outcomes will be obtained at follow-up appointments at 2, 6 and 12 months after randomisation. The primary outcomes are: pain, global perceived effect and patient-generated measure of disability at 2 months and recurrence at 12 months.

Discussion: This trial will be the first placebo-controlled trial of motor control exercise. The results will inform best practice for treating chronic low back pain and prevent its occurrence.

Background

The problem of chronic low back pain

Low back pain is the main cause of work absence and disability in industrialised societies. Approximately 10–20% of patients with low back pain develop chronic pain, defined as pain persisting for more than 3 months. Additional to their pain these patient's health problems typi-

cally include reduced physical function and psychological distress[1]. These patients use more than 80% of health care resources for back problems, and treatment has a low success rate [2].

In 2002, *arthritis and musculoskeletal disorders* were announced as the new National Health Priority Area in

recognition of the major health and economic burden that these diseases place on the Australian community [3]. Amongst this group of diseases back pain is both the most prevalent and most costly single disease [4]. The 2001 National Health Survey revealed that *chronic back pain* is the most prevalent illness from the seven National Health Priority Areas [5].

The severity of chronic pain can be described with four hierarchical grades, Grades I–IV, that consider the pain intensity and the degree of disability associated with the pain [6]. An Australian population-based survey, noted that 22% of respondents reported chronic pain with 39% of respondents classed as Grade I (least severe), 35% as Grade II, 14% as Grade III and 13% as Grade IV (most severe) [7]. The most common cause of chronic pain was low back pain (45% of cases).

Effectiveness of treatments for chronic low back pain

While there are a myriad of treatment options for chronic low back pain, there is only one clinical practice guideline for chronic non-specific low back pain: The European Guideline [8]. This guideline and the relevant Cochrane reviews [9] provide the most reliable sources of evidence on treatment for this condition. Unfortunately the Cochrane reviews provide fairly bleak reading for both clinicians and patients. Most of the reviews (7/13) concluded that the treatment under review was of unknown value. Five of the thirteen reviews concluded that there was some evidence for the treatment under review however significant limitations for each treatment were noted. These limitations included: no long term effect (e.g. back school); serious side effects (e.g. muscle relaxants); small effect size (e.g. massage); treatment improves outcomes other than pain (e.g. work conditioning) and no information available on patient or dose selection (e.g. behavioural treatment). The European Guideline produced similar conclusions [8]. In only one Cochrane review, the review of multidisciplinary rehabilitation/functional restoration, did the reviewers conclude that there was strong evidence for the therapy. However the reviewers also noted that these programs were only effective when they included >100 hours of therapy. Because these programs are multidisciplinary they are typically provided in a tertiary setting and because of the amount of time involved they are also very expensive. Accordingly functional restoration is usually reserved for the most severe cases of chronic low back pain.

The majority of patients with chronic low back pain has less severe pain (i.e. Grades I–III) and are typically managed in primary care. Not surprisingly clinicians find managing chronic low back pain difficult with qualitative research reporting that therapists' inability to identify effective treatment choices for their patients makes them

state clinicians perhaps feel 'helpless' 'disillusioned' and 'pessimistic' [10]. Studies of patients reveal similar negative feelings and emotions [11].

To address this major problem, we plan to begin a coordinated program of research in which treatments that seem most promising are rigorously evaluated in randomised controlled trials. We define 'most promising treatments' as those that (i) appear to have clinically important effects that are maintained in the long term, (ii) are readily available and of modest cost and (iii) there is biological plausibility for the effect. Exercise therapy is our first candidate for evaluation in this program of research because it satisfies each of these three criteria, however at present trials have reported conflicting results.

While some trials of exercise therapy have reported large, durable and clinically important effects of treatment [12,13] others have not [14]. The uncertainty is reflected in the conclusion of the Cochrane review of exercise therapy: '...there is conflicting evidence on the effectiveness of exercise therapy...' [15]

Many factors are likely to have contributed to the inconsistent results across trials. Importantly, interpretation of the results of exercise trials is difficult because most trials have been pragmatic trials, comparing two active treatments delivered in routine practice (e.g. *exercise* vs. *usual medical care* [12]; *exercise* vs. *physiotherapy* [16]) These comparisons cannot provide a clear estimate of the effects of exercise treatment because most of the comparison treatments are also of unknown efficacy. Secondly, there has been insufficient appreciation by researchers conducting trials and by reviewers summarising trials of the wide variety of forms exercise can take and also trials do not control the quality of exercise intervention. While exercise is typically regarded as a single class of treatment we believe that this level of conception is inappropriate and analogous to not distinguishing between different classes and doses of drugs when prescribing medication. The types of exercise programs for chronic low back pain vary widely from land-based exercise versus exercise in water to isolated trunk exercise versus a walking program and it is unlikely that all programs are equally effective for all patients. Lastly, methodological quality varies greatly across previous exercise trials, for example in the Cochrane review [15] the least sound trial attended to none of the nine methodological criteria while the best attended to seven of the nine. Because methodological quality has been shown to affect the results of trials in other areas of health care [17] it is likely that a lack of rigor has contributed to the inconsistent results.

It is not sensible to talk about evaluating the efficacy of exercise without specifying the type of exercise. We have

chosen to measure the efficacy of *motor control exercise* (sometimes called specific spinal stabilisation exercise) for chronic low back pain, rather than other forms of exercise, because it is a widely used form of exercise and there is an extensive body of literature that provides a rationale for the mechanism of action. The only way to clearly establish the value of motor control exercise in the management of chronic low back pain is to evaluate the efficacy of this form of exercise therapy in a methodologically sound randomised placebo-controlled trial. Prior to conducting a placebo-controlled trial of exercise we felt that it was prudent to identify the most promising form of exercise that would subsequently be evaluated in the placebo-controlled trial. To do this we conducted a randomised controlled trial where 160 patients were randomised to an 8 week program of either motor control exercise or general exercise[18].

The trial demonstrated that both programs were accompanied by large improvements in pain and disability. Motor control exercise produced significantly better outcomes in the short term, and there was a trend for motor control exercise to produce better outcomes at 6 month follow-up. Accordingly we have chosen to evaluate motor control exercise in the proposed trial. Our choice coincides with the research agenda set by the 2004 European Guideline: "The effectiveness of specific types of exercise therapy needs to be further evaluated. This includes the evaluation of spinal stabilisation exercises..." [8] p 7.

Motor control exercise: treatment rationale

The use of motor control exercise is based on research that has shown that:

- (i) People with low back pain have changes in the strategy for control of the trunk muscles in that activity of the deep muscles is impaired (delayed, less tonic) and these muscles are atrophied[19,20].
- (ii) Although all muscles contribute to control of movement and stability of the spine, the deep muscles have a critical role for control of intervertebral motion [21-25], but with the potential advantage of allowing dynamic control of the spine.
- (iii) Evidence that people with back pain tend to adopt a strategy for increased stiffness and stability at the expense of spinal function [26].
- (iv) Non-resolution of changes in the deep muscle system is linked to recurrences of low back pain [27].

The evidence above underpins the primary aim of motor control exercise, which is to re-establish normal control of the deep spinal muscles, reducing the activity of more

superficial muscles that tend to stiffen the spine and have increased activity in low back pain, and then maintain normal control during progressively more demanding physical and functional tasks[28].

The key feature of the motor control exercise approach is the training of the deep trunk muscles in isolation *before* progressing to demanding tasks that train coordination of the deep and the superficial trunk muscles [28]. However, unlike functional restoration approaches, training the deep trunk muscles in isolation from the superficial trunk muscles is difficult. In order to teach patients how to contract the deep muscles of the spine, in addition to clinical skills of palpation and observation [29] physiotherapists need to use technical devices such as pressure monitors, electromyography and ultrasound imaging to provide feedback to the patient.

The premise of the motor control approach is that simple functional exercise alone does not re-establish coordination of the trunk muscles. This premise is supported by the finding that the adaptation of these muscles to pain is still present following recovery from an episode of low back pain, when patients have returned to normal functional levels [19,20]. Furthermore, recent data confirm that coordination of the abdominal muscles can be restored with training of specific activation of the trunk muscles, but not a simple activation during a sit up task [30]. Notably, non-resolution of muscle dysfunction is associated with increased back pain recurrence [27]. Also, asymptomatic people with normal activity levels who are unable to perform a task that is thought to reflect voluntary activation of the deep trunk muscles, are ~6 times more likely to develop back pain than asymptomatic people who are able to perform the same task [31].

Motor control exercise: level I and II evidence

At present there is no evidence for the efficacy of motor control exercise in the treatment of chronic low back pain. No systematic review of motor control exercise has been published, although one is being completed by our group. While the majority of trials (5 of 8) report that motor control exercise is effective in the management of chronic or recurrent low back pain most (7 of 8) have permitted co-intervention so that the contribution of motor control exercise is unclear. Additionally, all of these previous trials have used other treatments of unknown efficacy as the comparison intervention and so treatment efficacy cannot be measured. For example the earliest trial [12] reported that motor control exercise is more effective than usual medical care however this result provides an ambiguous estimate of treatment effectiveness because other trials have reported that sham physiotherapy treatments are more effective than usual medical care [32].

We will evaluate the efficacy of motor control exercise in a placebo-controlled randomised controlled trial. The results of our study will be invaluable for more efficacious evidence-based management of patients with non-specific chronic low back pain. Once efficacy is established, we will be able to progress to measuring whether there are additive or multiplicative effects of other treatments that are commonly administered as co-interventions with motor control exercise and thus to being able to make valid recommendations for their use.

Methods

Overview of research design

The study will be a randomised, blinded, placebo-controlled trial of a motor control exercise program for patients with chronic low back pain. The exercise program will consist of 12 individually supervised half-hour sessions over an 8-week period with treatment outcomes measured at 2 months, 6 months and one year.

Hypotheses

(i) An 8-week motor control exercise program designed to restore control of the trunk muscles improves pain, disability and global perceived effect in participants with chronic low back pain at 2 months follow-up.

(ii) The improvements in pain, disability and global perceived effect following motor control exercise are maintained at 6 and 12 months follow-up.

(iii) At 12 month follow up recurrence is less in the motor control exercise group.

Subject recruitment

A total of 154 participants will be recruited into the study. Participants will be screened for suitability for motor control exercise according to usual clinical practice. The screening instruments identify participants who are unsuitable for exercise management of their low back pain because of significant co-morbidity (serious spinal pathology, contraindication to exercise). A clinical assessment will identify patients who we expect would best be managed by a motor control exercise program rather than some other form of exercise or physiotherapy management.

Screening

To screen for serious pathology, the physiotherapist will conduct a diagnostic triage [33]. Participants in whom serious spinal pathology is suspected will be excluded from the trial and referred to their medical practitioner for review. Potential participants will be screened for contraindications to exercise using the Physical Activity Readiness Questionnaire [34]. If a volunteer provides a positive response to items 1, 2, 3, 4, 6 or 7, the trial phys-

iotherapist will discuss the case with the referring medical practitioner and if necessary a medical review will be undertaken to exclude any contraindication to exercise as listed in the ACSM guidelines [34].

The clinical assessment used to ensure that the motor control approach is indicated is based on the key text [28] and is a normal part of clinical assessment of low back pain. The assessment involves evaluation of the motor control strategy during a specific trunk muscle task – drawing in of the lower abdomen while maintaining an isometric contraction of the medial back muscles. The following criteria constitute correct performance of the task:

1. Moderate and sustained activation (> 10 seconds) of transversus abdominis
2. Moderate and sustained activation (> 10 seconds) of the lumbar multifidus muscles
3. Little or no activation of the global trunk muscles
4. No spinal or rib cage movement.
5. Normal breathing

Evaluation of task performance including satisfaction of the above criteria is dependent on the clinical skills of the physiotherapist. Patients who are unable to perform this task correctly will be considered suitable for motor control exercise.

Participants will be included if they meet all of the following inclusion criteria:

- Non-specific low back pain +/- leg pain of at least 3 months duration
- Currently seeking care for low back pain
- Aged greater than 18 and less than 80 years
- Comprehends English
- Clinical assessment indicates that the subject is suitable for motor control exercise
- Expects to continue residing in SWSAH region for study duration.

Participants will be excluded if they have any of the following:

- Suspected or confirmed serious spinal pathology (fracture, metastatic, inflammatory or infective diseases of the

spine, cauda equina syndrome/widespread neurological disorder)

- Suspected or confirmed pregnancy
- Unable to speak English
- Nerve root compromise (2 of strength, reflex or sensation affected for same nerve root)
- Spinal surgery
- Scheduled for major surgery during treatment or follow-up period
- Any of the contraindications to exercise listed on page 42 of the ACSM guidelines [34]
- Any contraindication to pulsed ultrasound or pulsed shortwave.

Specific spinal pathology or contraindication to treatment may be suspected based on the results of the screening questionnaire and the Physical Activity Readiness Questionnaire. If the assessor suspects the presence of any pathology or contraindication to treatment, these subjects should be further investigated and medical clearance obtained, if necessary.

Assessment and allocation

Outcome measures

Measures of outcomes will be obtained at follow-up appointments at 2, 6 and 12 months after randomisation. To maximise attendance at these follow-ups, appointments will be made by phone and then a letter will be sent confirming appointment and a reminder phone call will be made 24 hrs before the appointment. Every attempt (within ethical constraints) will be made to obtain outcome data, regardless of subject's compliance with trial protocols. Follow-up measures will be scored by an investigator who is blinded to group allocation. At 2 months, information about side effects of treatment will be collected from all participants using open-ended questioning.

Following the screening consultation, personal characteristics (age, gender, ethnicity, religion, weight, height, level of education, employment status, doctor's details and contact information) and information about symptoms of low back pain will be collected (eg DASS 21 [35]; Chronic Pain Grade Questionnaire) The following treatment efficacy variables will be measured at baseline, 2, 6 and 12 months.

1. Average pain intensity over last week (0–10 scale) [36–38]
2. Patient-generated measure of disability (Patient-Specific Functional Scale) [36–38]
3. Global perceived effect (Global Perceived Effect Scale) [36–38]
4. Condition-specific measure of disability (Roland Morris Disability Questionnaire) [36–38]
5. Recurrence at 12 months

The primary outcomes are pain, GPE and PSFS at 2 months and recurrence at 12 months.

Randomisation

Participants will be allocated to treatment group using sealed opaque envelopes. The allocation sequence will be generated by author CM. Participants will be scheduled to receive their first treatment within one week of randomisation.

Interventions

Contemporary physiotherapy practice in exercise prescription is to assess each patient and to implement the form of exercise that is most relevant to the particular clinical presentation. At present this widely accepted approach relies primarily upon the clinical expertise of the therapist. We have elected to evaluate motor control exercise delivered in this manner because this approach is regarded as contemporary best practice.

Participants in each group will receive 12 half hour treatments over an 8-week period, i.e. 2 sessions/week in the first month and 1 session/week in the second month. The treatment sessions are designed to become less frequent over time to encourage independence and continuation of exercise when therapy is complete. This is consistent with current clinical practice.

The *motor control exercise* program is based on the treatment approach reported by O'Sullivan et al [12], Richardson et al [28], and Moseley [39]. A brief description is provided below.

At the first session, participants will be comprehensively assessed and then will be prescribed exercises aimed at improving function of specific muscles of the low back region to be conducted in sessions 2–11. Stage 1 involves the most commonly prescribed exercise aimed at retraining multifidus (a back muscle) and transversus abdominus (a deep abdominal muscle); these exercises will be supplemented with exercises for the pelvic floor muscles,

breathing control and control of spinal posture. Participants will be taught how to contract these muscles independently from the superficial trunk muscles [28,40]. Physiotherapists will use real-time ultrasound biofeedback to enhance learning of the tasks. When participants are able to perform these exercises, they will be gradually upgraded until the patient is able to maintain isolated contractions of these muscles for 10 seconds, up to a maximum of 10 repetitions, during normal respiration [28]. When this level of competence has been achieved, patients will be considered ready to progress to Stage 2.

Stage 2 of the approach involves increasing the complexity of the exercise by progressing through a range of functional tasks and exercises targeting coordination of trunk and limb movement and maintenance of trunk stability. The range and progression of exercises is well set-out in clinical texts [28] and is individualised to the patient based on this presentation. Participants require the ongoing support of a trained physiotherapist to ensure correct performance of the exercises. Session 12 is a discharge session where the patient's progress will be reviewed and patients will be prescribed exercises to continue at home.

The *placebo* intervention is 20 minutes of detuned short wave diathermy and 5 minutes of detuned ultrasound for 12 sessions over an eight week period. This attention control will be used because there is no known treatment effect from the detuned machines, but it has been established in previous trials (including one of our own [37]) that participants view this as a credible treatment. To increase the perceived credibility of the attention control, participants will undergo an examination including routine screening for contraindications at the first consultation and the normal clinical reassessment that would occur with the active forms of these interventions at each subsequent treatment. Each placebo treatment session will be 30 minutes in duration to match the active treatment sessions.

Participants in both treatment groups will be asked not to seek other treatments for their chronic low back pain and where possible not to change current medications during the treatment period. Several mechanisms will be used to ensure that the trial protocol is consistently applied. Protocol manuals will be developed and staff will be trained to ensure that screening, assessment, randomisation and treatment procedures are conducted according to protocol. To ensure standardisation across sites we will hold regular meetings with site visits and teleconferencing. An independent researcher will monitor a randomly chosen subset to ensure adherence to assessment, randomisation and treatment procedures.

If a participant is concerned about his or her condition during the study, the physiotherapist will screen for potentially serious pathology and, where appropriate, refer the patient to a medical practitioner. The medical practitioner will be asked not to request the participant's group allocation unless it is deemed necessary for medical care. At the completion of the exercise program, patients will be encouraged to continue the home exercise routine demonstrated at the discharge session. Participants will be free to seek other treatment after the experimental period.

After the first treatment session the patient will complete a treatment credibility scale [41]. At the 8 week follow-up information about side-effects of treatment will be collected using open-ended questioning. At the 12 month follow-up the participants will be asked to rate the helpfulness, understanding and friendliness of therapist and helpfulness of treatment and to nominate which treatment they thought they received. Additionally information about other treatment received for their low back pain during the study period will be sought using open-ended questioning.

Data integrity

The integrity of trial data will be monitored by regularly scrutinising data sheets for omissions and errors. Data will be double entered and the source of any inconsistencies will be explored and resolved.

Data analysis

Treatment efficacy

In our primary analysis, we will use a regression model to test for the effect of treatment on outcome at 2, 6 and 12 months follow-up with the baseline value of the outcome entered as a covariate. A treatment effect size will be calculated for each of the follow-up time points and, if there is a statistically significant treatment effect at any time point, we will also calculate number needed to treat (NNT) to achieve pain recovery (pain < 1 out of 10: [42]) and 95% CI. The recurrence outcome will be analysed with logistic regression.

Predictor of response to treatment

We will include an interaction term baseline DASS-21 depression score \times group to the regression analysis to see if the effect of motor control exercise is influenced by the baseline DASS-21 depression score.

Sample size calculations

We have designed the study to detect a clinically important difference of 1 unit on the 0–10 pain intensity scale (estimate for SD = 2.00), 1 unit on the 0–10 patient specific functional scale (estimate for SD = 1.8); 1 unit on the 0–10 Global Perceived Effect Scale (estimate for SD = 1.65) and 4 units on the 24 item Roland Morris Disability

Questionnaire (estimate for SD = 4.9). We have taken the SD estimates from a trial we completed that recruited a similar patient cohort [37]. With specifications of alpha = 0.05, power = 0.80 a sample size of 77 participants per group is required to detect an effect size of 0.50 SD (the smallest effect size we have specified for the four outcomes). Based on the results of the same trial [37] we have allowed for 15% non-compliance with treatment, 15% loss to follow-up, and assumed a correlation between baseline and change scores of outcomes of 0.5. Accordingly we will recruit 77 participants per group or 154 participants in total.

Justification of study design

Placebo

Designing an appropriate placebo treatment that mimics a physiotherapy exercise program is challenging. The sham interventions used in previous exercise trials do not satisfy the criteria of being both inert (e.g. the use of hot packs) and credible (e.g. allocation to a treatment waiting list). Accordingly, we will use sham electrotherapy as a control. This sham is clearly inert and is regarded as a credible treatment by participants. [37] To ensure that participants remain unaware of study group, it is necessary to carefully describe the study to patients. In the previous trial where we used sham electrotherapy as a control for exercise, we used the following description:

'In this trial normal physiotherapy treatment and placebo physiotherapy treatment will be provided. A placebo treatment is a harmless treatment delivered at less than the effective dose. We will not tell you which type of treatment you will receive and it is unlikely that you could distinguish them.'

Trial staff described the placebo intervention as 'pulsed ultrasound' and 'pulsed shortwave' and explained to patients that they would probably not feel any sensation during treatment.

Controlling bias

The trial has been designed to include key methodological features that have been recognised as minimising bias in clinical trials. These features include: true randomisation, concealed allocation, specification of eligibility criteria, blind outcome assessment, patient blinding, blind analysis and intention-to-treat analysis. The nature of the treatments precludes blinding of treatment provider. Trial staff will be trained to ensure consistency of screening, assessment, randomisation and treatment procedures. Participant's perception of the credibility of treatment will be determined after the first treatment [41]; and at the completion of treatment both assessors and participants will be asked to identify what treatment they think the participant received.

Outcomes

Measures of pain symptoms, disability and generic health status will be taken from the 'core set' of outcome measures for clinical research recently advocated by an international panel of back pain researchers [43]. The panel considered factors such as reliability, validity and responsiveness before recommending a measure. We have supplemented the back-related disability measure advocated in the core set (Roland Morris) with a patient-generated measure of disability (Patient-Specific Functional Scale) because there is evidence that patient-generated measures of disability are more responsive than condition-specific measures [37,44].

Conclusion

We have presented the rationale and design of a randomized controlled clinical trial evaluating the effect of motor control exercise versus placebo in patients with chronic LBP. The results of this trial will be published as soon as they are available.

Competing interests

All author(s) declare that they have no competing interests.

Authors' contributions

CGM, JL, PWH, KMR, GLM, RDH and LOPC were responsible for the design of the study. LOPC and JM will act as trial coordinators. All authors have read and approved the final manuscript.

References

1. Bogduk N: Management of chronic low back pain. *Medical Journal of Australia* 2004, **180**(2):79-83.
2. Waddell G: **The clinical course of low back pain.** In *The Back Pain Revolution* Edinburg, Churchill Livingstone; 1998.
3. Commonwealth Department of Health and Ageing: **National Health Priorities and Quality.** 2003 [<http://www.health.gov.au/pq/index.htm>].
4. Australian Institute of Health and Welfare: **Australia's health 2000: the seventh biennial health report of the Australian Institute of Health and Welfare.** Canberra, AIHW; 2000.
5. Australian Bureau of Statistics: **National Health Survey. Summary of Results.** Canberra, Australian Bureau of Statistics; 2002.
6. Von Korff M, Ormel J, Keefe FJ, Dworkin SF: **Grading the severity of chronic pain.** *Pain* 1992, **50**(2):133-149.
7. Blyth F, March L, Cousins M: **Chronic pain-related disability and use of analgesia and health services in Sydney community.** *Medical journal of Australia* 2003, **179**:84-87.
8. Cost B13 working group: **European Guidelines for the Management of Chronic Non-Specific Low back Pain.** 2004.
9. Cochrane Collaboration Back Review Group: **Reviews.** 2004 [<http://www.cochrane.iwh.on.ca/review.htm>].
10. Chew CA, May CR: **The benefits of back pain.** *Family Practice* 1997, **14**(6):461-465.
11. Hansson M, Bostrom C, Harms-Ringdahl K: **Living with spine-related pain in a changing society--a qualitative study.** *Disability & Rehabilitation* 2001, **23**(7):286-295.
12. O'Sullivan P, Twomey L, Allison G: **Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis.** *Spine* 1997, **22**:2959-2967.
13. Kankaanpaa M, Taimela S, Airaksinen O, Hanninen O: **The efficacy of active rehabilitation in chronic low back pain. Effect on**

- pain intensity, self-experienced disability, and lumbar fatigability.** *Spine* 1999, **24(10)**:1034-1042.
14. Yelland MJ, Glasziou PP, Bogduk N, Schluter PJ, McKernon M: **Prolotherapy injections, saline injections, and exercises for chronic low-back pain: a randomized trial.** *Spine* 2004, **29(1)**:9-16.
 15. van Tulder M, Malmivaara A, Esmail R, Koes B: **Exercise therapy for low back pain: a systematic review within the framework of the cochrane collaboration back review group.** *Spine* 2000, **25(21)**:2784-2796.
 16. Hemmila HM, Keinänen-Kiukaanniemi SM, Levoska S, Puska P: **Long-term effectiveness of bone-setting, light exercise therapy, and physiotherapy for prolonged back pain: a randomized controlled trial.** *Journal of Manipulative & Physiological Therapeutics* 2002, **25(2)**:99-104.
 17. Egger M, Juni P, Bartlett C, Hohenstein F, Sterne J: **How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study.** *Health Technology Assessment* 2003, **7(1)**:1-76.
 18. Ferreira PH: **Effectiveness of specific stabilisation exercises for chronic low back pain.** In *Physiotherapy*, The University of Sydney; 2004.
 19. Hodges PW, Richardson CA: **Inefficient muscular stabilization of the lumbar spine associated with low back pain. A motor control evaluation of transversus abdominis.** *Spine* 1996, **21(22)**:2640-2650.
 20. Hides JA, Stokes MJ, Saide M, Jull GA, Cooper DH: **Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain.** *Spine* 1994, **19(2)**:165-172.
 21. Hodges P, Kaigle Holm A, Holm S, Ekstrom L, Cresswell A, Hansson T, Thorstensson A: **Intervertebral stiffness of the spine is increased by evoked contraction of transversus abdominis and the diaphragm: in vivo porcine studies.** *Spine* 2003, **28(23)**:2594-2601.
 22. Wilke HJ, Wolf S, Claes LE, Arand M, Wiesend A: **Stability increase of the lumbar spine with different muscle groups. A biomechanical in vitro study.** *Spine* 1995, **20(2)**:192-198.
 23. Solomonow M, Zhou BH, Harris M, Lu Y, Baratta RV: **The ligamento-muscular stabilizing system of the spine.** *Spine* 1998, **23(23)**:2552-2562.
 24. Macintosh J, Bogduk N: **The biomechanics of the lumbar multifidus.** *Clinical Biomechanics* 1986, **1**:205-213.
 25. Kaigle AM, Holm SH, Hansson TH: **Experimental instability in the lumbar spine.** *Spine* 1995, **20(4)**:421-430.
 26. van Dieen JH, Selen LP, Cholowicki J: **Trunk muscle activation in low back pain patients, an analysis of literature.** *Journal of Electromyography Kinesiology* 2003, **13**:333-351.
 27. Hides JA, Jull GA, Richardson CA: **Long-term effects of specific stabilizing exercises for first-episode low back pain.** *Spine* 2001, **26(11)**:E243-8.
 28. Richardson CA, Jull GA, Hodges PW: **Therapeutic Exercise for Spinal Segmental Stabilization in Low Back Pain.** Edited by: Livingstone C. Edinburgh; 1999.
 29. Urquhart DM, Hodges P, Allen TJ, Story IH: **Abdominal muscles recruitment during a range of voluntary exercises.** *Manual Therapy* 2005, **10**:144-153.
 30. Tsao H, Hodges P: **Specific abdominal retraining alters motor coordination in people with persistent low back pain: ; Sydney.** ; 2005.
 31. Moseley GL: **Impaired trunk muscle function in sub-acute neck pain: etiologic in the subsequent development of low back pain?** *Manual Therapy* 2004, **9(3)**:157-163.
 32. Koes BW, Bouter LM, van Mameren H, Essers AH, Verstegen GM, Hofhuizen DM, Houben JP, Knipschild PG: **The effectiveness of manual therapy, physiotherapy, and treatment by the general practitioner for nonspecific back and neck complaints. A randomized clinical trial.[see comment].** *Spine* 1992, **17(1)**:28-35.
 33. Waddell G, Feder G, McIntosh A, Lewis M, Hutchison A: **Low back pain evidence review.** London, Royal College of General Practitioners; 1996.
 34. ACSM: **ACSM's Guidelines for Exercise Testing and Prescription.** Baltimore, Williams and Wilkins; 1995.
 35. Henry JD, Crawford JR: **The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample.** *British Journal of Clinical Psychology* 2005, **44**:227-239.
 36. Scrimshaw SV, Maher C: **Responsiveness of visual analogue and McGill pain scale measures.** *Journal of Manipulative & Physiological Therapeutics* 2001, **24(8)**:501-504.
 37. Pengel L: **Outcome of recent onset low back pain (PhD Thesis).** In *Physiotherapy* Sydney, The University of Sydney; 2004.
 38. Maher C, Latimer J, Refshauge K: **Atlas of Clinical Tests and Measures for Low Back Pain.** Melbourne, Australian Physiotherapy Association; 2001.
 39. Moseley L: **Combined physiotherapy and education is efficacious for chronic low back pain.** *Australian Journal of Physiotherapy* 2002, **48(4)**:297-302.
 40. Twomey L, Taylor J: **Physical therapy for low back pain. Clinics in Physical Therapy.** Edited by: Livingstone C. New York, ; 1994.
 41. Borkovec T, Nau S: **Credibility of analogue therapy rationales.** *Journal of Behaviour Therapy and Experimental Psychiatry* 1972, **3**:257-260.
 42. McGuirk B, King W, Govind J, Lowry J, Bogduk N: **Safety, efficacy, and cost effectiveness of evidence-based guidelines for the management of acute low back pain in primary care.** *Spine* 2001, **26(23)**:2615-2622.
 43. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korf M, Waddell G: **Outcome measures for low back pain research. A proposal for standardized use.** *Spine* 1998, **23(18)**:2003-2013.
 44. Stratford P, Gill C, Westaway M, Binkley J: **Assessing disability and change on individual patients: a report of a patient specific measure.** *Physiotherapy Canada* 1995, **47**:258-263.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2474/6/54/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Chapter Three

Exercise for chronic low back pain: A randomised, placebo-controlled trial


Chapter Three is published as:

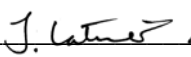
Costa LOP, Maher CG, Latimer J, Hodges PW, Herbert RD, Refshauge KM, McAuley JH, Jennings MD (2009) Exercise for chronic low back pain: A randomised, placebo-controlled trial. Submitted for publication to *PLoS Medicine* on 16.03.09.

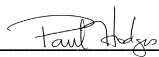
**Statement from co-authors confirming authorship contribution of the
PhD candidate**

As co-authors of the paper “Exercise for chronic low back pain: A randomised, placebo-controlled trial”, we confirm that Leonardo Costa has made the following contributions:


- Conception and design of the research
- Data collection
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

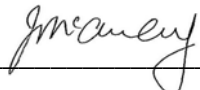
Christopher G Maher _____  _____ Date 20.03.2009

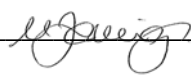
Jane Latimer _____  _____ Date 20.03.2009

Paul W Hodges _____  _____ Date 20.03.2009

Kathryn Refshauge _____ Date 20.03.2009

Robert D Herbert _____  _____ Date 20.03.2009

James McAuley _____  _____ Date 20.03.2009

Matthew D Jennings _____  _____ Date 20.03.2009

**Exercise for chronic low back pain:
A randomised, placebo-controlled trial.**

Leonardo OP Costa¹, Christopher G. Maher¹, Jane Latimer¹,
Paul W. Hodges², Robert D. Herbert¹, Kathryn M. Refshauge³, James H. McAuley¹
and Matthew D. Jennings⁴.

¹Musculoskeletal Division, The George Institute For International Health, Sydney, Australia. ²NHMRC Centre of Clinical Research Excellence in Spinal Pain, Injury and Health, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia. ³Faculty of Health Sciences, The University of Sydney, Australia. ⁴Liverpool Hospital, Sydney South West and Western Sydney Area Health Services, Sydney, Australia.

Address: Musculoskeletal Division, The George Institute For International Health, PO Box M201, Missenden Rd Sydney, NSW 2050 Australia. Email: lc03060@gmail.com

Financial support information: This study was funded by a Research & Development grant from the University of Sydney and by the Physiotherapy Research Foundation. Leonardo Costa is a PhD student supported by CAPES – Ministério da Educação – Brazil and Pontifícia Universidade Católica de Minas Gerais – Brazil; Chris Maher, Paul Hodges and Rob Herbert hold research fellowships funded by the National Health and Medical Research Council of Australia.

Background: The evidence that exercise is effective for treatment of chronic low back pain comes from trials that are not placebo-controlled. We aimed to investigate the efficacy of motor control exercise for chronic low back pain. The trial was registered with the Australian Clinical Trials Registry, ACTRN012605000262606 and the protocol was prospectively published.

Methods and findings: In total 154 patients with chronic low back pain of >12 weeks duration were randomly assigned to 12 sessions of exercise or placebo over 8 weeks. Measures of outcomes were obtained at baseline and at follow up appointments 2, 6 and 12 months after randomisation. Primary outcomes were pain, function, and the patient's global impression of recovery measured at 2 months. Of the 154 participants randomised to groups, 152 attended the 2-month follow up (98.7%) and 145 attended both 6 and 12-month follow up (94.2%). The exercise intervention improved function and patient's global impression of recovery, but did not clearly reduce pain, at 2 months. The mean effect of exercise on function was 1.1 points (95%CI, 0.3 to 1.8), the mean effect on global impression of recovery was 1.5 points (95%CI, 0.4 to 2.5) and the mean effect on pain was 0.9 points (95%CI, -0.01 to 1.8), all measured on 11 point scales. Secondary outcomes also favoured motor control exercise. The main limitation was that clinicians could not be blinded to the intervention they provided.

Conclusions: Motor control exercise produces small short-term improvements in global impression of recovery and function, but not pain, for people with chronic low back pain. This pattern of results was similar at 6 and 12 months follow-up.

Introduction

Low back pain is a major health and socioeconomic problem and is associated with high costs in care, work absenteeism and disability worldwide[1,2]. A recent inception cohort study demonstrated that 43% of patients with acute low back pain presenting to primary care developed chronic low back pain and nearly a third of them did not recover within one year[3].

Exercise is endorsed as an effective treatment for chronic low back pain in most clinical practice guidelines[2,4,5]. However at present there are no placebo controlled trials of exercise for chronic low back pain[6,7]. The positive recommendations in guidelines are instead derived from trials comparing exercise to usual care[8,9], or to a wait list[10], or to no treatment[11]. These trials do not control for placebo effects and potentially provide biased estimates of the effect of exercise because they do not control for changes in patient and assessor behaviour caused by knowledge of treatment allocation[12,13].

In order to establish the efficacy of exercise for chronic low back pain we conducted the first placebo-controlled trial of motor control exercise (also known as specific stabilisation exercise). We chose to study motor control exercise as it is a popular form of exercise for spinal pain and there is a biologically plausible rationale for its efficacy[14,15,16].

Methods

Patients. The trial was conducted in an outpatient physiotherapy department of a university teaching hospital in Sydney, Australia. Consecutive patients seeking care

for chronic low back pain were screened for eligibility. To be eligible for inclusion participants had to have non-specific low back pain which is defined as pain and discomfort, localised below the costal margin and above the inferior gluteal folds, with or without referred leg pain of at least 3 months duration, be currently seeking care for low back pain, be aged between 18 and 80 years, comprehend English, and expect to continue residing in the study region for the study duration. In addition potential subjects underwent a simple trunk muscle test to determine that motor control exercise treatment was indicated[17,18]. Exclusion criteria were suspected or confirmed spinal pathology (e.g. tumor, infection, fracture, inflammatory disease), pregnancy, nerve root compromise, previous spinal surgery, scheduled for major surgery during treatment or follow up period, presence of any contraindication to exercise[19] or ultrasound or short-wave therapy. The study protocol was approved by the University of Sydney Human Research Ethics Committee.

Procedures. The randomisation sequence was computer-generated by one of the investigators who was not involved in recruitment. The sequence was blocked (block sizes of 4, 6 and 8, in random order). Allocation was concealed in sequentially numbered, sealed, opaque envelopes. Eligible patients were allocated to treatment groups by opening the next-numbered envelope.

Participants in each group received 12 half-hour treatments over an 8-week period (2 sessions/week in the first month and 1 session/week in the second month). The placebo treatment was designed to be structurally equivalent[20] to the active intervention, providing similar contact time with the physiotherapist. Both interventions were provided by three senior physiotherapists who received training

from experts in motor control exercise and placebo interventions. This training included a one-day workshop prior to the commencement of the study and 3 half-day follow up sessions during the trial period. Random audits and regular meetings provided by the same experts were conducted during the trial to monitor delivery of interventions.

The *motor control exercise* program was based on the treatment approach described in previous publications[9,21,22,23]. At the first session participants were comprehensively assessed by the physiotherapist and then were prescribed exercises that were individualised, based on the participant's presentation. The exercises were designed to improve function of specific muscles of the low back region and the control of posture and movement.

Stage 1 of the exercise program involved retraining of the multifidus (a back muscle) and transversus abdominus (a deep abdominal muscle). These exercises were supplemented with exercises for the pelvic floor muscles, breathing control and control of spinal posture and movement. The specific muscles that were trained depended on the initial assessment. Participants were taught how to contract these muscles independently from the superficial trunk muscles[24,25]. Physiotherapists used real-time ultrasound biofeedback to enhance learning of the tasks. The exercises were progressed until the patient was able to maintain isolated contractions of the target muscles for 10 repetitions of 10 seconds each, whilst maintaining normal respiration[24]. When this level of competence was achieved, patients were considered ready to progress to Stage 2.

Stage 2 of the exercise program involved increasing the complexity of the exercise by progressing through a range of functional tasks and exercises targeting coordination of trunk and limb movement, maintenance of optimal trunk stability, and improvement of posture and movement patterns. Participants required the ongoing support of a trained physiotherapist to ensure correct performance of the exercises. The participants were instructed to perform a set of home exercises to be performed daily. These exercises should be performed at the same level, in the same position as those demonstrated during the treatment session. Session 12 was a discharge session in which the patient's progress was reviewed and exercises were prescribed to be continued at home.

The *placebo* intervention consisted of 20 minutes of detuned short-wave diathermy and 5 minutes of detuned ultrasound for 12 sessions over an eight-week period. This form of placebo was used because the detuned machines do not provide a specific treatment effect, but it has been established in previous trials[26,27,28] that participants view this intervention as credible. To ensure the perceived credibility of the placebo intervention, physiotherapists followed the usual clinical routine for the delivery of the active form of these two treatments (e.g., by checking for contraindications, monitoring changes in symptoms, adjusting the detuned devices and appearing to progress the treatment). Each placebo treatment session lasted 30 minutes to match the duration of active treatment sessions.

A careful explanation was provided to patients to ensure they remained blinded to treatment allocation. We used the following description for the patients: *'In this trial normal physiotherapy treatment and placebo physiotherapy treatment will be*

provided. A placebo treatment is a harmless treatment delivered at less than the effective dose. We will not tell you which type of treatment you will receive and it is unlikely that you could distinguish them'. The trial staff described the placebo intervention as 'pulsed ultrasound and pulsed shortwave' and explained to patients that they would probably not feel any sensation during treatment. The active forms of these treatments delivered in pulsed mode do not produce heat; thus, previous experience with the treatments would not un-blind participants. The sham units were identical to active units (for example, the on and off lights illuminated and the output dial moved) except that they did not provide output. The nature of the interventions precluded blinding of the treatment provider.

Measures of outcomes were obtained at baseline and at follow up appointments 2, 6 and 12 months after randomisation. Primary outcomes were nominated in the trial protocol[18]. The primary outcomes were pain intensity over the last week (0-10 Numeric Rating Scale)[29], function (0-10 Patient-Specific Functional Scale)[30], and global impression of recovery (-5 to +5 Global Perceived Effect Scale) at 2 months and recurrence at 12 months[31]. Secondary outcomes were pain, function and patient's global impression of recovery measured at 6 and 12 months, and disability (0-24 Roland Morris Disability Questionnaire)[32] measured at 2, 6 and 12 months. Participants reported their outcomes by phone interview to an investigator who was blinded to the treatment allocation. Patients were asked not to discuss any aspect of their treatment with the assessor.

We also measured recovery and recurrence at 12 months. Patients were considered to have recovered if they reported that they had become pain-free and this pain-free

period lasted for at least one month[33]. Recurrences could only occur in patients who had recovered. Recurrence was defined as a new episode of low back pain that persisted for more than 24 hours[33,34].

Baseline data were collected prior to randomisation. The baseline data included all outcome measures and the participant's characteristics (age, gender, ethnicity, religion, weight, height, level of education and employment status). In addition we collected information about depressive symptoms (DASS-21)[35,36] to test if the effect of the exercise intervention on primary outcomes was influenced by the DASS-21 depression score.

Participants rated treatment credibility (0-24 treatment credibility scale)[37] after the first treatment session. They were asked about side effects at 2 months using open-ended questions[38]. At 12 months, patients were asked about treatment satisfaction, measured with a 4-item scale with questions about the therapist (i.e. how helpful, friendly and understanding the physiotherapist was) and about the treatment helpfulness in general. At 12 months patients were asked "*which treatment did you receive? Real physiotherapy treatment? Or a sham or pretend treatment?*" to check participant blinding.

Statistical Analysis. A sample size of 154 participants was nominated in the trial protocol[18]. We allowed for 15% non-compliance with treatment and 15% loss to follow-up, and assumed a correlation of 0.5 between baseline scores and outcomes. This sample size provides 80% power to detect an effect of exercise of 1 unit on the 0–10 pain intensity scale (estimated SD = 2.0), 1 unit on the 0–10 Patient Specific

Functional Scale (estimated SD = 1.8); 1 unit on the Global Perceived Effect Scale (estimated SD =1.7) and 4 units on the 24 item Roland Morris Disability Questionnaire (estimated SD = 4.9), when alpha is set at 0.05.

Data were double-entered. The statistical analysis was performed on an intention-to-treat basis. The statistician was given coded data, so was blinded to which group received the exercise intervention.

The mean effects of intervention on pain intensity, function, disability and global perceived effect were calculated using linear mixed models (random intercepts and fixed coefficients) which incorporated terms for treatment, time and the treatment by time interactions. The effect of time was non-linear so time was dummy coded and analysed as a categorical variable (i.e. three dummy variables were created for the categories 2, 6 and 12 months). The coefficients of the treatment by time interactions provided estimates of the effects of the exercise intervention. To determine whether baseline depression scores modified the effect of exercise, a secondary analysis was conducted in which a higher-level interaction term (baseline DASS-21 depression score \times group \times time) was added to each of the regression models.

As very few patients recovered according to our definition of being pain-free for 30 days during the study period, only a small subset of participants could experience a recurrence. To provide a measure relevant to all participants, we created a new outcome called 'persistent low back pain' which was coded as 'no' for participants who recovered AND did not have a recurrent episode within 12 months and 'yes' for all other participants. This outcome was tested in a *post-hoc* analysis. We calculated

confidence intervals for the risk difference using Newcombe's method based on Wilson's score method, without continuity correction[39].

Mixed models analyses were performed with Stata 9 (Stata Corp LP, Texas, USA). Other analyses were performed with SPSS 16.0 for windows (SPSS Inc, Chicago, USA). The study was prospectively registered with the Australian Clinical Trials Registry (ACTRN012605000262606) and the protocol published[18].

Role of funding sources. The study was funded by a Research & Development grant from the University of Sydney and the Physiotherapy Research Foundation. The funding sources had no role in study design, data collection, data analysis, interpretation of data, or writing of the trial report. The investigators had final responsibility in the decision to submit the report for publication.

Results

In total 220 participants seeking care for low back pain were screened for eligibility between October 2005 and December 2007 (figure 1). Seventeen patients chose not to participate and 49 were considered ineligible. The reasons for ineligibility were nerve root compromise (n=9), previous spinal surgery (n=8), serious spinal pathology (n=6), non-English speaker (n=6), scheduled for major treatment or surgery during the follow up period (n=5), low back pain of less than 12 weeks duration (n=7), aged older than 80 (n=1), contraindication to exercise (n=1), unable to commit to attend the treatment sessions due to distance (n=1) and advice from the trial therapists that the patient was not suitable for motor control exercise treatment (n=5) (for reasons of bilateral knee replacement, substance abuse, recent epilepsy

collapse, vascular claudication or Erdheim-Chester disease). Results of the simple trunk muscle task indicated that motor control exercise was suitable for all tested subjects and therefore no exclusions were performed based upon this criterion. Of the 154 participants randomised to groups, 152 attended the 2-month follow up (98.7%) and 145 attended both 6 and 12-month follow up (94.2%). The characteristics of the participants in the two groups were similar at baseline (table 1).

From 12 planned treatment sessions, the participants in the exercise group attended a mean of 8.8 sessions (SD 3.5) compared to 9.6 sessions (SD 3.0) for patients allocated to the placebo intervention. Most of the participants believed that they were allocated to a “real/active” intervention (85% of patients from the exercise group versus 84% of patients from the placebo group). The ratings of treatment satisfaction were similar in both groups with the medians ranging from 4 to 6 points (on a 0-6 point scale) (table 2).

Five patients (2 from the placebo group and 3 from the exercise group) reported mild adverse effects of the interventions. All adverse effects were temporary exacerbations of pain. None of the patients withdrew from the trial due to adverse effects. In total 10 patients from the exercise group and 14 patients from the placebo group reported co-interventions during the study period.

The exercise intervention improved function and the patient’s global impressions of recovery (table 3 and figure 2). At 2 months, exercise improved function by a mean of 1.1 points (95%CI, 1.8 to 0.3) on the 0-10 Patient Specific Functional Scale and also improved patient’s global impression of recovery by 1.5 points (95%CI, 2.5 to

0.4). There was not a clear effect of exercise on pain intensity at 2 months (-0.9 point (95% CI, -1.8 to 0.0) P=0.053) or 6 months, but there was a statistically significant effect at 12 months (1.0 point (95% CI, 0.1 to 1.9) in favour of the exercise group. During the study period few patients had become pain-free (recovered): 22% of the patients in the exercise group and 9% in the placebo group recovered. Ten percent of the exercise group and 7% of the placebo group recovered but then experienced a recurrence within 12 months. Consequently 88% of the exercise group and 98% of the placebo group had persistent pain at 12 months, although at a lower intensity for the exercise group (absolute risk reduction 10%; 95% CI, 1% to 19%; number need to treat 10).

Exercise improved disability at 2 months (-2.7 points (95% CI, -4.4 to -0.9) and 6 months (-2.2 points (95% CI, -4.0 to -0.5) but the differences were smaller and no longer significant at 12 months (difference = 1.0 point (95% CI, -0.8 to 2.8 points). Finally there was no evidence that depression was a predictor of response to treatment at 2 months for pain intensity, (β = -0.03, 95% CI, -0.10 to 0.04); global impression of recovery, (β = -0.05, 95% CI, -0.0. to 0.13) or function, (β = 0.10, 95% CI, -0.07 to 0.27).

Discussion

This is the first randomised, placebo-controlled trial of exercise for chronic low back pain. We found evidence of a beneficial, though small, effect of motor control exercise on global impression of recovery, function, disability at 2 months and risk of persistent pain at 12 months, but not pain intensity at 2 and 6 months and disability at 12 months. Most of the effects observed at short-term follow-up were maintained 12

months after randomisation. We also found that the effect of motor control exercise was not influenced by the level of depressive symptoms.

Our interpretation of the trial results is that exercise produces small clinical improvements but complete recovery is unlikely. Some patients and clinicians may not consider these effects clinically worthwhile. The effects are smaller than benchmarks for clinically important effects suggested by expert researchers in the low back pain field[40,41] and in recent clinical practice guidelines[2]. However we acknowledge that consensus has not been reached on this issue amongst back pain researchers; and study of patients[42] reveals an even wider range of views on how big an improvement in outcomes needs to be before it is considered worthwhile. Given this diversity of views clinicians may need to spend some time with patients considering exercise treatment outlining the likely outcomes and assisting the patient to decide if they wish to pursue the treatment.

The mean effects of exercise treatment were smaller than has been previously reported in some previous trials[6] however these trials include features associated with exaggerated treatment effects, such as lack of patient blinding and absence of controlling for placebo effects[12]. Our use of a placebo-controlled design provides control of potentially important sources of bias, so the effects of treatment we observed are less likely to be exaggerated than the effects observed in non-placebo controlled trials[12].

The exact biological basis for the efficacy of motor control exercise in patients with low back pain is still unclear[43] but if subjects can be taught to control their trunk

muscles while performing functional activities, then this may explain the improvements seen in function, disability and global impression of recovery[14,15]. There is some evidence that this training can change trunk muscle behaviour during functional tasks[44,45]. A range of mechanisms have been proposed to explain the effect of motor control training on pain. These include reduced load and improved quality of movement[46] as a result of improved coordination of trunk muscles. Such changes in control may be mediated by plastic changes at the motor cortex or elsewhere in the motor system[47].

Our study demonstrated that motor control exercise produced a small reduction in the risk of persistent pain at 12 months. This finding is supported by earlier work[14] that suggests that patients who have continuing impairment of the deep trunk muscles experience more recurrent low back pain episodes. This provides a rationale for why those in the exercise group, who retrained the deep trunk muscles, experienced less recurrence than those in the placebo group who had no such training.

While systematic reviews of the efficacy of exercise for chronic low back pain[6] have generally concluded that exercise is effective, most reviews also signal some uncertainty in their conclusions because of methodological concerns in the available trials. Our trial avoided the main methodological problems of previous trials by using a placebo control and blinding patients and assessors. In addition the trial was prospectively registered and the trial protocol was published[18]. Lastly we took steps to ensure treatment quality by using experienced clinicians who were trained to

deliver the treatments according to the protocol, and we monitored treatment delivery.

The main limitation of our study was that the trial therapists were not blinded to the treatment allocation. Blinding of therapists is impossible in placebo-controlled trials of exercise. We tried to minimize the effect of unblinding by training the trial therapists to provide a credible placebo treatment and by auditing placebo treatment sessions. We believe that these steps were effective because scores on credibility and treatment satisfaction were similar in both treatment groups. Nevertheless we cannot exclude the possibility that the lack of therapist blinding introduced some degree of bias into our results.

Although it could be argued that our choice of placebo was not perfect, we believe that this choice was the best possible. We do not know of a “placebo exercise” that is both credible and inert. This problem is not unique to the study of exercise and similar problems with developing an appropriate placebo were found in trials of complex non-pharmaceutical interventions such as spinal manipulative therapy[28,48] or acupuncture[49]. Our selection of sham electrotherapy as a placebo was primarily based upon the knowledge that these machines do not share the same specific components of the exercise intervention and also because they have been used successfully in previous randomised controlled trials[26,28].

Our study provides evidence that motor control exercise is better than placebo in patients with chronic low back pain. Most of the effects observed at short term were maintained at 6 and 12 months follow-up. Our results suggest that this intervention

should be considered for patients with chronic low back pain in order to improve disability, function, and global impression of recovery, and to improve pain intensity in the long term, but not the short term.

References

1. Cost B13 working group (2004) European Guidelines for the Management of Chronic Non-Specific Low back Pain.
2. Chou R, Qaseem A, Snow V, Casey D, Cross JT, et al. (2007) Diagnosis and treatment of low back pain: A joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 147: 478-491.
3. Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, et al. (2008) Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ* 337: 1-7.
4. Cost B13 working group (2006) European Guidelines for the Management of Chronic Non-Specific Low back Pain. *Eur Spine J* 15 S192-S300.
5. Chou R, Huffman LH (2007) Nonpharmacologic therapies for acute and chronic low back pain: A review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 147: 492-504.
6. Hayden JA, van Tulder MV, Malmivaara A, Koes BW (2005) Exercise therapy for treatment of non-specific low back pain. *Cochrane Database Syst Rev*: CD000335.
7. Machado LAC, Kamper S, Herbert RD, Maher CG, McAuley JH (2008) Analgesic effects of treatments for non-specific low back pain: a meta-analysis of placebo-controlled randomized trials. *Rheumatology* doi: 10.1093/rheumatology/ken470.
8. O'Sullivan PB, Twomey LT, Allison GT (1997) Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis. *Spine* 22: 2959-2967.
9. Moseley L (2002) Combined physiotherapy and education is efficacious for chronic low back pain. *Aust J Physiother* 48: 297-302.
10. Risch SV, Norvell NK, Pollock ML, Risch ED, Langer H, et al. (1993) Lumbar strengthening in chronic low-back-pain patients - physiological and psychological benefits. *Spine* 18: 232-238.

11. Kuukkanen T, Mälkiä E, Kautiainen H, Pohjolainen T (2007) Effectiveness of a home exercise programme in low back pain: a randomized five-year follow-up study. *Physiother Res Int* 12: 213-224.
12. Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical-evidence of bias - dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273: 408-412.
13. Juni P, Altman DG, Egger M (2001) Systematic reviews in health care - Assessing the quality of controlled clinical trials. *BMJ* 323: 42-46.
14. Hides JA, Jull GA, Richardson CA (2001) Long-term effects of specific stabilizing exercises for first-episode low back pain. *Spine* 26: E243-248.
15. Hides JA, Stokes MJ, Saide M, Jull GA, Cooper DH (1994) Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain. *Spine* 19: 165-172.
16. Hodges PW, Richardson CA (1996) Inefficient muscular stabilization of the lumbar spine associated with low back pain - A motor control evaluation of transversus abdominis. *Spine* 21: 2640-2650.
17. Ferreira PH (2004) Effectiveness of specific stabilisation exercises for chronic low back pain. Sydney: The University of Sydney.
18. Maher CG, Latimer J, Hodges PW, Refshauge KM, Moseley GL, et al. (2005) The effect of motor control exercises versus placebo in patients with chronic low back pain. *BMC Musculoskelet Disord* 6: 54.
19. ACSM (1995) ACSM's Guidelines for exercise testing and prescription. Baltimore: Williams and Wilkins.
20. Machado LAC, Kamper SJ, Herbert RD, Maher CG, McAuley JH (2008) Imperfect placebos are common in low back pain trials: a systematic review of the literature. *Eur Spine J* 17: 889-904.
21. O'Sullivan P, Twomey L, Allison G (1997) Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis. *Spine* 22: 2959-2967.

22. Richardson CA, Jull GA, Hodges PW, Hides JA (1999) Therapeutic Exercise for Spinal Segmental Stabilization in Low Back Pain. Edinburg: Churchill Livingstone.
23. Hodges PW, Ferreira PH, Ferreira ML (2007) Lumbar spine: Treatment of instability and disorders of movement control. In: Magee DJ, Zachazewski JE, Quillen WS, editors. Scientific Foundations and Principles of Practice in Musculoskeletal Rehabilitation: Saunders.
24. Richardson CA, Jull GA, Hodges PW (1999) Therapeutic Exercise for Spinal Segmental Stabilization in Low Back Pain; Livingstone C. Edinburg.
25. Twomey L, Taylor J (1994) Physical therapy for low back pain. Clinics in Physical Therapy.; Livingstone C. New York.
26. Pengel LHM, Refshauge KM, Maher CG, Nicholas MK, Herbert RD, et al. (2007) Physiotherapist-directed exercise, advice or both for subacute low back pain: a randomized trial. *Ann Intern Med* 146: 787-796.
27. Ferreira ML, Ferreira PH, Latimer J, Herbert RD, Hodges PW, et al. (2007) Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. *Pain* 131: 31-37.
28. Hancock MJ, Maher CG, Latimer J, McLachlan AJ, Cooper CW, et al. (2007) Assessment of diclofenac or spinal manipulative therapy, or both, in addition to recommended first-line treatment for acute low back pain: a randomised controlled trial. *Lancet* 370: 1638-1643.
29. Scrimshaw SV, Maher C (2001) Responsiveness of visual analogue and McGill pain scale measures. *J Manipulative Physiol Ther* 24: 501-504.
30. Stratford PW, Gill C, Westaway M, Binkley JM (1995) Assessing disability and change on individual patients: a report of a patient-specific measure. *Physiother Can* 47: 258-263.
31. Pengel LHM, Refshauge KM, Maher CG (2004) Responsiveness of pain, disability and physical impairment outcomes in patients with low back pain. *Spine* 29: 879-883.
32. Roland M, Morris R (1983) A study of natural history of back pain, part 1: development of a reliable and sensitive measure of disability in low back pain. *Spine* 8: 145-150.

33. de Vet HCW, Heymans MW, Dunn KM, Pope DP, van der Beek AJ, et al. (2002) Episodes of low back pain - A proposal for uniform definitions to be used in research. *Spine* 27: 2409-2416.
34. Stanton TR, Henschke N, Maher CG, Refshauge KM, Latimer J, et al. (2008) After an episode of acute low back pain, recurrence is unpredictable and not as common as previously thought. *Spine* 33: 2923-2928.
35. Henry JD, Crawford JR (2005) The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol* 44: 227-239.
36. Haggman S, Maher CG, Refshauge KM (2004) Screening for symptoms of depression by physical therapists managing low back pain. *Phys Ther* 84: 1157-1166.
37. Borkovec TD, Nau S (1972) Credibility of analogue therapy rationales. *J Behav Ther Exp Psych* 3: 257-260.
38. Bent S, Padula A, Avins AL (2006) Better ways to question patients about adverse medical events: A randomized, controlled trial. *Ann Intern Med* 144: 257-261.
39. Newcombe RG (1998) Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 17: 873-890.
40. van Tulder M, Malmivaara A, Hayden JA, Koes B (2007) Statistical significance versus clinical importance. *Trials on exercise therapy for chronic low back pain as example. Spine* 32: 1785-1790.
41. Ostelo R, Deyo RA, Stratford P, Waddell G, Croft P, et al. (2008) Interpreting change scores for pain and functional status in low back pain - Towards international consensus regarding minimal important change. *Spine* 33: 90-94.
42. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL (2000) Defining the clinically important difference in pain outcome measures. *Pain* 88: 287-294.
43. Hodges P (2008) Transversus abdominis: a different view of the elephant. *Br J Sports Med* 42: 941-944.

44. Tsao H, Hodges P. Specific abdominal retraining alters motor coordination in people with persistent low back pain; 2005; Sydney.
45. Tsao H, Hodges PW (2007) Immediate changes in feedforward postural adjustments following voluntary motor training. *Exp Brain Res* 181: 537-546.
46. Hodges PW, Moseley GL (2003) Pain and motor control of the lumbopelvic region: effect and possible mechanisms. *J Electromyogr Kinesiol* 13: 361-370.
47. Tsao H, Galea M, Hodges P. Skilled motor training induces reorganisation of the motor cortex and is associated with improved postural control in chronic low back pain; 2008; Glasgow, UK. International Association for the Study of Pain.
48. Hancock MJ, Maher CG, Latimer J, McAuley JH (2006) Selecting an appropriate placebo trial of spinal manipulative therapy. *Aust J Physiother* 52: 135-138.
49. Paterson C, Dieppe P (2005) Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ* 330: 1202-1205.

Table 1. Baseline characteristics

Characteristic	Exercise Group (n=77)	Placebo Group (n=77)
Mean age, (SD), years	54.6 (13.0)	52.8 (12.7)
Female gender (%)	45 (58)	48 (62)
Mean low back pain duration (SD), weeks	334.8 (392.3)	328.2 (395.1)
Mean height (SD), meters	1.65 (0.09)	1.64 (0.10)
Mean weight (SD), kilos	74.5 (17.5)	75.9 (15.3)
Smoker (%)	21 (27)	19 (25)
Taking analgesics (%)	61 (79)	58 (75)
Participating in moderate exercise (%)*	41 (53)	51 (66)
Work Status (%)		
Working full-time	6 (8)	13 (17)
Working part-time	5 (7)	3 (4)
Not working	20 (26)	12 (16)
Not seeking employment	46 (60)	49 (64)
Education (%)		
School certificate	19 (25)	17 (22)
High school certificate	19 (25)	18 (23)
Trade certificate	5 (6)	9 (12)
Diploma	4 (5)	5 (7)
Advanced diploma	0	1 (1)
Bachelor degree	11(14)	12 (16)
Postgraduate degree	4 (5)	0
Other (lower than school certificate)	15 (20)	15 (20)
General health status (%)		
Excellent	3 (4)	8 (10)
Very good	18 (23)	12 (16)
Good	38 (49)	44 (57)
Fair	14 (18)	7 (9)
Poor	4 (5)	6 (8)
Mean Depression Anxiety Stress Scale (SD)		
Depression §§	11.4 (12.9)	11.2 (13.4)
Anxiety	11.9 (11.1)	11.8 (12.2)
Stress ¶¶¶	14.1(11.8)	14.4 (12.5)
Mean primary outcome scores (SD)		
Pain intensity ¶¶	6.8 (2.1)	6.6 (2.0)
Global Perceived Effect ††	-1.9 (2.5)	-2.1 (2.4)
Patient-Specific Functional Scale **	3.3 (1.7)	3.3 (1.8)
Mean secondary outcome scores (SD)		
Roland Morris Disability Questionnaire ‡‡	13.1 (5.0)	13.4 (4.9)

* Moderate exercise was any type of exercise in a moderate intensity greater than 30 minutes duration at least three times per week; §§ 0 (no depression) to 42 (high depression); |||| 0 (no anxiety) to 42 (high anxiety); ¶¶¶ 0 (no stress) to 42 (high stress); ¶¶ 0 (no pain) to 10 (worst pain possible); †† -5 (vastly worse) to 5 (completely recovered), with 0 being unchanged; ** 0 (cannot perform activity) to 10 (can perform activity at pre-injury level); ‡‡ 0 (no disability) to 24 (high disability).

Table 2. Credibility and treatment evaluation comparisons

Characteristic	Exercise Group (n=77)	Placebo Group (n=77)
Median credibility scale (IQR)		
How confident do you feel that this treatment can help relieve your pain? ^{¶¶}	5 (2)	4 (2)
How confident do you feel that this treatment will help you manage your pain? ^{¶¶}	5 (2)	4 (3)
How confident would you be in recommending this treatment to a friend who suffered from similar complaints? ^{¶¶}	5 (2)	4 (3)
How logical does this treatment seem to you? ^{**}	5 (2)	4 (3)
Median treatment evaluation (IQR)		
Therapist helpfulness ^{††}	5 (2)	5 (2)
Therapist understanding ^{‡‡}	6 (1)	6 (1)
Therapist friendliness ^{§§}	6 (0)	6 (0)
Treatment helpfulness ^{††}	4 (2)	4 (3)

^{¶¶} 0 (not at all confident) to 6 (absolutely confident).

^{**} 0 (not at all logical) to 6 (very logical).

^{††} 0 (not at all helpful) to 6 (extremely helpful)

^{‡‡} 0 (not at all understanding) to 6 (extremely understanding)

^{§§} 0 (not at all friendly) to 6 (extremely friendly)

Table 3. Effects of exercise and placebo

Variable	Unadjusted mean outcome (SD)		Exercise versus Placebo	
	<i>Exercise</i>	<i>Placebo</i>	Relative change (95% confidence intervals)	P value
Pain¶				
2 months	4.6 (2.8)	5.6 (2.6)	-0.9 (-1.8 to 0.0)	0.053
6 months	5.0 (2.9)	5.6 (2.5)	-0.5 (-1.4 to 0.5)	0.335
12 months	5.0 (2.9)	6.3 (2.3)	-1.0 (-0.1 to -1.9)	0.030
Global Perceived Effect††				
2 months	1.3 (3.2)	0.0 (3.1)	1.5 (2.5 to 0.4)	0.005
6 months	1.5 (2.6)	0.3 (3.0)	1.4 (2.4 to 0.3)	0.010
12 months	1.2 (2.7)	-0.3 (2.9)	1.6 (2.6 to 0.6)	0.003
Function**				
2 months	5.2 (2.4)	4.1 (2.3)	1.1 (1.8 to 0.3)	0.004
6 months	5.3 (2.7)	4.3 (2.6)	1.0 (1.8 to 0.3)	0.007
12 months	5.5 (2.6)	4.0 (2.6)	1.5 (2.2 to 0.7)	<0.001
Disability‡‡				
2 months	9.6 (6.5)	11.9 (5.9)	-2.7 (-4.4 to -0.9)	0.003
6 months	10.3 (7.0)	12.2 (6.7)	-2.2 (-4.0 to -0.5)	0.014
12 months	11.4 (7.8)	12.3 (6.4)	-1.0 (-2.8 to 0.8)	0.271

¶ Measured with a 0 (no pain) to 10 (worst pain possible) numerical rating scale.

†† -5 (vastly worse) to 5 (completely recovered), with 0 being unchanged.

** Measured with Patient-Specific Functional Scale with participant selecting three activities and rating their ability to perform the activity on a 0 (cannot perform activity) to 10 (can perform activity at pre-injury level). Summary score is the mean of the three activities.

‡‡ Measured with 24 item Roland Morris Disability Questionnaire. Scale range 0 (no disability) to 24 (high disability).

Primary outcomes are highlighted in grey.

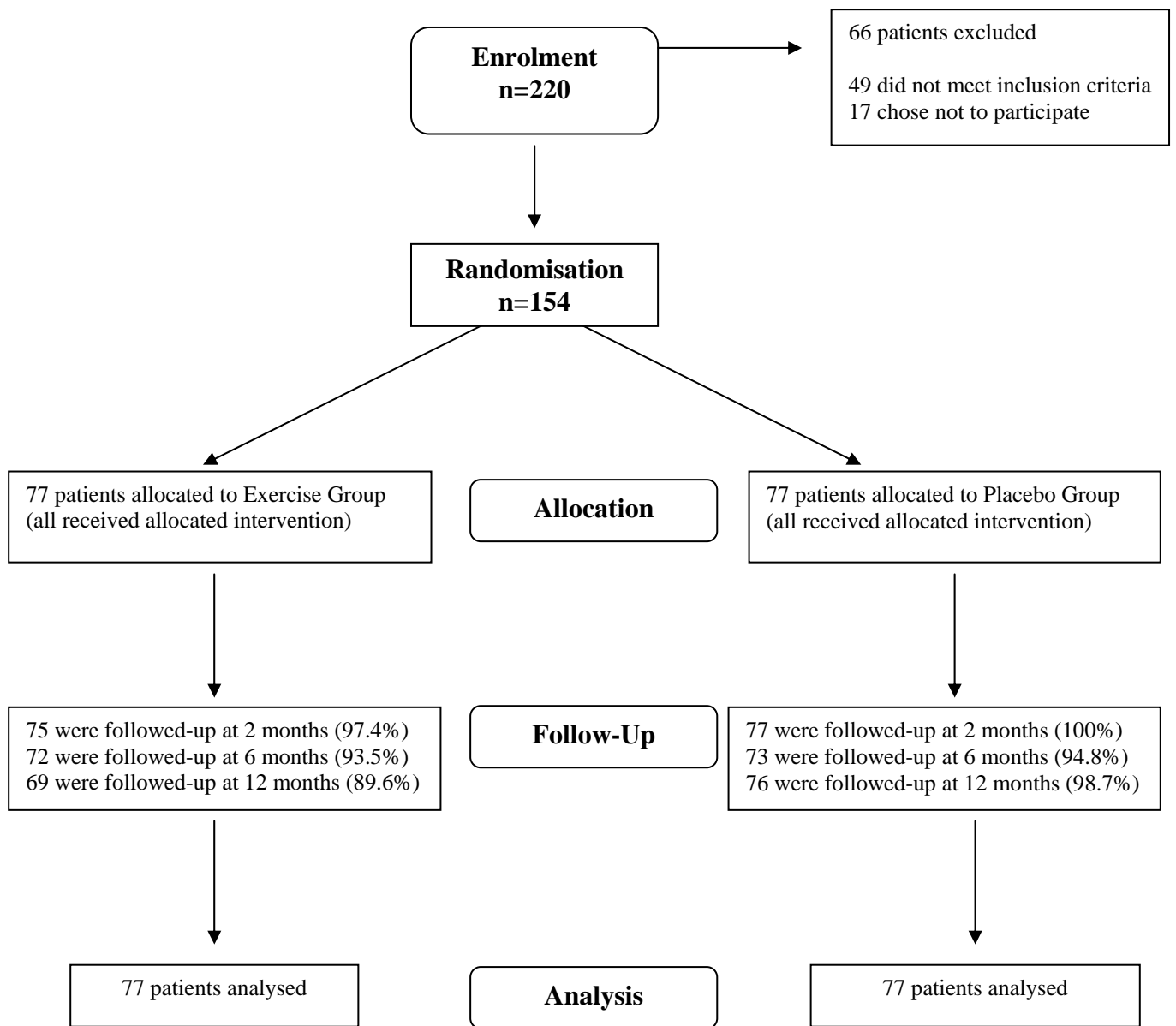


Figure 1. Study flow diagram

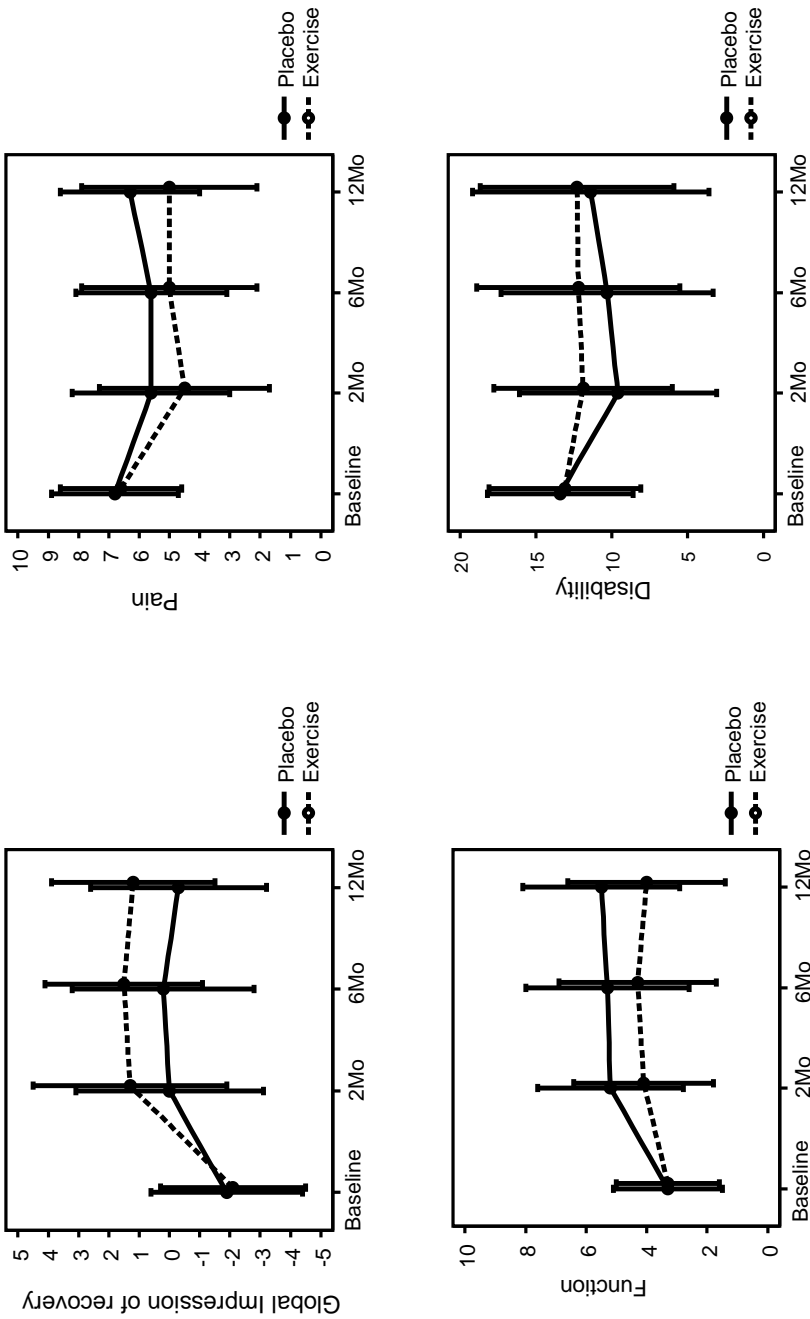


Figure 2. Outcomes in the two treatment groups

Values shown are unadjusted means (SDs). Measurements were obtained at baseline, 2, 6 and 12 months, but the data are slightly offset in the figure for clarity. Higher scores represent better outcomes for global impression of recovery and disability, and lower scores represent better outcomes for pain and function.

Chapter Four

Reproducibility of rehabilitative ultrasound imaging for the measurement of abdominal muscle activity: a systematic review

Chapter Four is published as:

Costa LOP, Maher CG, Latimer J, Smeets RJEM (2009) Reproducibility of real-time ultrasound for the measurement of abdominal muscle activity: a systematic review.

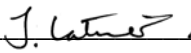
Accepted for publication in *Physical Therapy* on 02.02.09.

**Statement from co-authors confirming authorship contribution of the
PhD candidate**

As co-authors of the paper “Reproducibility of real-time ultrasound for the measurement of abdominal muscle activity: a systematic review”, we confirm that Leonardo Costa has made the following contributions:

- Conception and design of the research
- Data collection
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

Christopher G Maher _____  _____ Date 20.03.2009

Jane Latimer _____  _____ Date 20.03.2009

Rob JEM Smeets _____  _____ Date 20.03.2009

**Reproducibility of rehabilitative ultrasound imaging for the measurement of
abdominal muscle activity: a systematic review.**

**Leonardo Oliveira Pena Costa*⁺, Chris G. Maher*, Jane Latimer*,
Rob J.E.M. Smeets*.**

*Musculoskeletal division, The George Institute for International Health & Faculty
of Medicine – The University of Sydney, Australia.

PO Box M201 Missenden Road Sydney NSW 2050 Australia.

Phone: 612 82382440

Email: lc3060@gmail.com

⁺Corresponding author

Sources of support: Leonardo Costa is a PhD student supported by CAPES –
Ministério da Educação – Brazil and Pontifícia Universidade Católica de Minas
Gerais – Brazil; Chris Maher holds a research fellowship funded by the National
Health and Medical Research Council of Australia.

Abstract

Background and purpose: Rehabilitative ultrasound imaging (RUSI) measures of abdominal wall muscles are used to indirectly measure muscle activity. These measures are used to identify suitable patients, and to monitor progress of motor control exercise treatment of low back pain. The purpose of this study was to systematically review reproducibility studies of RUSI for measuring thickness of abdominal wall muscles.

Methods: Eligible studies were identified via searches of electronic databases. We also searched personal files and tracked references of the retrieved studies via the Web of Science Index. Two independent reviewers extracted data and assessed methodological quality. Due to heterogeneity of the studies' designs, pooling the data for a meta-analysis was not possible.

Results: 21 studies were included and these studies were typically of low quality and studied healthy subjects rather than people seeking care for low back pain. The studies reported good to excellent reliability for single measures of *thickness* and poor to good reliability for measures of *thickness change* (reflecting the muscle activity). Interestingly, no studies checked reliability of measures of *the difference in thickness changes over time* (representing improvement or deterioration in muscle activity).

Conclusions and Discussion: The current evidence of the reproducibility of RUSI for measuring abdominal muscle activity is mainly based upon studies with suboptimal designs and the study of healthy subjects. The critical question of whether RUSI provides reliable measures of improvement in abdominal muscle activity remains to be evaluated.

Introduction

The use of motor control exercise in the treatment of non-specific low back pain (LBP) has become increasingly popular in clinical practice. The rationale for the use of motor control exercise arises from the view that the activity of the deep abdominal muscles is critical for the dynamic control of the lumbar spine, with poor control resulting in lumbar spine symptoms^{1,2}. Most of the studies^{3,4} to date that have measured the activity of the deep spinal muscles are based on fine-wire electromyography (EMG) examination which is costly, uncomfortable and has associated risks such as infection, making its use in clinical practice difficult. An alternate approach is to indirectly measure the recruitment of the abdominal muscles by assessing morphologic changes (thickness changes) using rehabilitative ultrasound imaging (RUSI)^{5,6}. There is some evidence that RUSI measures of thickness change are correlated to EMG measures of muscle activity at low levels of contraction force (up to approximately 30% of maximal)⁷⁻⁹. This has resulted in the increasing use by healthcare professionals of ultrasound machines to assess motor control deficits and also to provide feedback for patients receiving treatment for low back pain.

It is widely accepted that clinical measurements need to be reproducible. Without an acceptable level of reproducibility, the clinical utility of assessment tools becomes substantially compromised¹⁰. It is important to state that reproducibility (i.e. degree to which repeated measurements in stable persons (test-retest) provide similar answers¹¹) should be understood as an umbrella term for reliability and agreement¹², where reliability could be defined as “the extent to which patients can be distinguished from each other, despite measurement errors (relative measurement

error)” and agreement defined as “the extent to which the scores on repeated measures are close to each other (absolute measurement error)”¹¹. In an ideal scenario, a reproducibility study should be designed with particular attention to 5 points¹¹; firstly a reproducibility study should be performed in patients with the condition in which the test will be used (e.g. ultrasound tests for the abdominal wall in patients with low back pain); secondly, the evaluation of reproducibility should be performed in a manner as similar as possible to the conditions used in clinical practice; thirdly, the study must be controlled for the order of the tests and for memory bias (which can be easily performed with blinding, an appropriate time interval and with randomisation /counterbalancing procedures for ordering of the tests); fourthly, the study must be sufficiently statistically powered, and finally the study must be analysed in a way that the results could be reasonably generalized to a certain population of clinicians (i.e. appropriate description of the tester and appropriate statistical analysis)¹³.

In the last decade a large number of studies evaluating the reproducibility of RUSI measures of abdominal muscle activity have been published¹⁴⁻¹⁷ and some of these studies have been reviewed in a non-systematic design¹⁸. To date there is no comprehensive systematic review that has attempted to investigate the reproducibility of RUSI measurements of the activity of the abdominal wall muscles. The objective of this study was to systematically review all reproducibility studies of RUSI for abdominal wall muscles.

Methods

Data sources and searches

Studies were identified through searches of MEDLINE (1950 to 2008), EMBASE (1974 to 2008) and CINAHL (1982 to 2008). The results of the searches were combined in an *Endnote X* software file. Additionally, hand searches of journals, references lists and textbooks related to ultrasound imaging were performed. We also searched personal files and tracked references of the retrieved studies via the Web of Science Index.

There were no language restrictions. A record of the number of papers retrieved and the number of papers included was kept. The search terms are displayed in table 1.

Study selection

To be included a study had to meet two criteria: firstly the study had to involve any type of reliability and or agreement of any type of ultrasound measurements (B or M mode) for any of the abdominal wall muscles and, secondly the characteristics of the participants must be described (i.e. healthy individuals/ LBP patients etc.). Relevant studies were identified by one of the authors (LOPC) and admitted to the study with agreement from a second author (RS).

Data extraction

Data from eligible studies were extracted by two independent reviewers (LOPC and RS). Table 2 presents all items that were extracted from the studies.

The reliability/agreement indexes were extracted for 3 different measures: *thickness* (i.e. static measures of muscle thickness at rest or contracted); *thickness changes* (i.e.

measuring the muscle activity by determining the degree of change in thickness between the resting and contracted states); or *differences in thickness changes over time* (i.e. measuring improvements/deterioration of muscle activity described above). We consider *thickness changes* and *differences in thickness changes* to be the most important measures, since these reflect use of the measure in current clinical practice.

The studies were also divided according to the study design into three categories: 1) studies reporting the reproducibility of taking repeated measurements of the same set of images, 2) studies reporting the reproducibility of repeating the total measurement procedure i.e. positioning the participant, positioning the ultrasound transducer, acquiring the images and measuring the images, and 3) studies reporting the reproducibility of a portion of the whole procedure (for example keeping the patient in the same position, but repositioning the transducer and acquiring new images for analysis).

Quality Assessment

The quality of the studies was rated, using item number 6 (reproducibility) from the *Quality Criteria for Measurement Properties*¹¹ (table 3), by two independent reviewers (LOPC and RS). This item evaluates the design of the study, as well as the reproducibility values by analysing two dimensions: reliability and agreement. These criteria form a checklist that considers both the methodological quality of the reproducibility testing and the results from the testing and so is somewhat different from scales used to measure the methodological quality of clinical trials¹⁹.

Results

From the search strategy, 315 potentially relevant studies were found. From these only 21 studies were considered eligible for data analysis (figure 1), being 17 full manuscripts from peer-reviewed journals and 4 abstracts from conference presentations²⁰⁻²³. Twelve studies calculated the reproducibility of the whole process of measuring (i.e. re-positioning the participant and ultrasound transducer, obtaining the images and measuring them)^{8,14-17,20,22,24-28} with ICC values ranging from 0.81-0.92 for static images and 0.26-0.85 for *thickness changes*. Of the studies that did not evaluate the whole protocol six studies calculated the reproducibility of measures from the same images only^{22,23,27-30} with ICC values ranging from 0.62-0.99 (static images) and 0.48-0.78 (*thickness changes*), two studies partially repeated the process of positioning the participant/transducer, but fully repeated the process of obtaining the images and measured them^{31,32} with ICC values ranging from 0.81-0.92 (static images only, no values for *thickness changes* was found), and finally we were unable to classify four studies due to unclear reporting^{21,33-35}. The ICC values arose from different ICC models and therefore caution should be taken in interpreting these results.

Given the heterogeneity of the studies in terms of design and statistical analysis, pooling of the data for a meta-analysis approach was not possible; therefore the presentation of our results is descriptive.

Table 4 describes the characteristics of the eligible studies. There were large differences in the sample sizes used in the reproducibility tests, ranging from 8 to 70 participants. Remarkably, only two studies recruited patients seeking care for their

LBP, 3 studies recruited a mixed sample of asymptomatic volunteers and LBP patients and 16 studies recruited only normal subjects for the tests. Information about the assessors, blinding and how the order of the tests was controlled for bias, were only presented in 4, 1 and 4 studies respectively. Sixteen studies used B mode ultrasound for collecting the images while the Transversus Abdominis was the most commonly investigated muscle (20 of 21 studies). Finally the time interval between the tests ranged from “immediately after” to 7 days; and 13 studies investigated the reproducibility while performing a voluntary task.

Table 5 describes the results of the eligible studies, although all studies investigated the reliability of measures of abdominal muscle *thickness*, only 6 investigated the reliability of *thickness changes* (reflecting muscle activity) and none tested the reproducibility of the *differences in thickness changes over time* (reflecting improvement / deterioration in muscle activity). The ICC was the reliability coefficient used in 18 studies however a range of forms of this statistic were used. There were 5 studies that used the ICC_{1,k}, 2 studies the ICC_{2,k}; 6 studies the ICC_{3,k} and 4 studies did not specify which type of ICC was chosen. Unfortunately only 5 studies provided confidence intervals for the ICC. Some studies used Pearson’s *r* and coefficient of variation as a measure of reliability. Agreement was calculated in 12 studies (being 12 for *thickness* and 3 for *thickness changes*) most of the studies used the SEM as the agreement parameter; additionally 2 studies used Bland and Altman Plots and one calculated the Minimum Detectable Change (MDC).

In terms of reliability while more than 80% of the ICC values for measuring *thickness* ranged from 0.80 to 1.00, most ICC values for measuring *changes in thickness* were

less than 0.70. Interestingly the ICC values tend to be slightly lower in the 5 studies that used participants with LBP, compared with the 16 studies that recruited asymptomatic participants only (table 5).

Table 6 presents the quality assessment of the 21 studies summarising each criterion as positive, doubtful, negative or no information. None of the studies demonstrated positive ratings for both reliability and agreement. For the reliability of *thickness* no studies were rated as positive, 20 as doubtful and 1 as negative. For the reliability of *thickness change* no study out of 6 was rated as positive, 2 as negative and 4 as doubtful. Twelve studies provided some information in regards to agreement (SEM or Bland and Altman plots or Smallest Detectable Change), with all of them rated as doubtful. The major reasons for doubtful/negative ratings were the lack of precise information about the time interval between the measures, lack of precise information about the ICC type, lack of statistical power (i.e. less than 50 subjects for the analysis)³⁶ and ICC values below 0.70.

Discussion

This review highlights the limitations of existing research evaluating the reproducibility of RUSI measures of abdominal wall muscles. Few studies analysed the reproducibility for the measurement of *thickness change* and no studies evaluated the reproducibility of the *difference in thickness change over time*. The available studies were frequently of low quality, recruited healthy subjects and evaluated only a portion of the RUSI measurement protocol. The existing data is of limited value in estimating the reproducibility of RUSI measures undertaken in a clinical setting to guide a motor control exercise program for low back pain.

The whole process of performing ultrasound measures has multiple potential sources of error (e.g. accuracy of measurements of distance, identification of landmarks, ability to perform the tasks properly, and position of patient/transducer). Additionally it has to be acknowledged that trial to trial variation in performance of the activation tasks is expected. It would be useful to consider whether modifications in the test protocol may enhance the reproducibility of US measures (especially for *thickness changes*). One approach that has been shown to enhance reliability of other low back assessments is to further standardise the protocol³⁷⁻³⁹.

The main methodological weaknesses found in the studies can be summarised into four main issues: (1) Generalisability of findings (due to sampling issues and due to the description of the assessor); (2) inadequate statistics; (3) bias (due to absence of blinding and/or due to not controlling the order of the tests); and (4) the lack of studies that investigated the reliability and especially the agreement of *thickness changes* and *differences in thickness changes over time*.

The generalisability of the findings from the individual studies selected from this systematic review is substantially limited given the fact that from 21 studies, only 2 recruited patients with low back pain^{22,28} and 3 recruited a mixed sample of LBP patients and healthy individuals^{16,23,25}. Additionally only 4 studies provided some description/source of the assessors^{22,27,28,35}. Moreover, from the studies cited above, only two provided the description of the assessor and recruited patients seeking care for LBP^{22,28}. We believe that clinically relevant studies must recruit participants seeking care for the condition in which the test will be used, and also include a

description of the assessor to enable better understanding of their qualifications, skills and length of training for future clinical comparisons.

The most widely used statistical test for the calculation of reliability was the Intraclass Correlation Coefficient (ICC) (18 of 21 occasions) which is a recommended option for testing reliability for continuous scales⁴⁰ (which is the case in studies of muscle thickness). There are multiple types of ICC and the choice of the correct ICC model depends on 3 considerations^{40,41}: 1) the wish to generalise, or not, the findings to other assessors 2) whether the same set of assessors rate each subject 3) if the authors are interested in the reliability of an individual assessor or the reliability of the mean rating of a group of assessors. Often, in the case of clinical research, only one judge is used and also it is important to generalise the results, therefore ICC type 2 (i.e ICC_{2,1}) or type 1 (i.e ICC_{1,1}) are preferred. ICC_{3,1} or ICC_{3,k} (k=number of assessors) should only be used if the authors do not want to generalise their results⁴². We found 4 studies that provided no information on the ICC type^{8,20,31,33} and 6 studies used ICC type 3 only as their reliability index^{16,23,25,27-29}. Pearson's r was used as a measure of reliability in one study²¹ and the coefficient of variation (CV) was used also in one study.³⁴ However the use of Pearson's r and CV are likely to provide overly optimistic estimates of reliability as they do not consider the 'between judges variance'^{40,41}. We considered that investigators should use ICC_{2,1} or ICC_{1,1} as a measure of reliability for *thickness* or *thickness changes* as they should provide the most relevant estimation of reliability. Agreement was analysed by 12 studies (10 studies calculated the Standard Error of the Measurement – SEM and 2 used Bland and Altman Plots) and the evidence for all studies was classified as doubtful due to small sample sizes and small time interval between the measures.

An issue that needs to be borne in mind is that many of the studies report the reproducibility of the mean of replicate measures. While this is an accepted method of enhancing the reliability of a measure it does make the measurement protocol more time consuming. In a similar fashion it needs to be remembered that some studies use highly trained raters, intricate equipment to control the position of the participant and load cells to standardise the activation of trunk muscles. While each of these three elements makes sense, they again limit the generalisability of the results.

Surprisingly we found that most of the studies calculated the reproducibility of *thickness* only. The problem with this approach is that in a clinical setting, the most important measures would be either *thickness changes* (comparing one image in rest state to another during muscle activity) which was calculated in 6 studies^{16,20,22,24,25,35} or *differences in thickness changes over time* (comparing thickness changes in two different time points for quantifying improvement/deterioration) which was not performed in any study. We suggest strongly that more studies investigating the reproducibility of *thickness changes* and *differences in thickness changes over time* should be undertaken.

We believe that our study provides important information for clinicians and researchers about the use of RUSI for abdominal wall muscles. It is important for clinicians to understand the limited evidence for reproducibility of the measures made over time when used to document success of a motor control treatment program. Additionally researchers have to acknowledge that the most important clinical questions about reproducibility of RUSI for abdominal wall muscles have been not answered and further studies are urgently needed.

References

1. O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: Maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 2005;10:242-55.
2. Hodges PW, Richardson CA. Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Exp Brain Res* 1997;114:362-70.
3. Hodges PW, Richardson CA. Altered trunk muscle recruitment in people with low back pain with upper limb movement at different speeds. *Arch Phys Med Rehabil* 1999;80:1005-12.
4. Tsao H, Hodges PW. Persistence of improvements in postural strategies following motor control training in people with recurrent low back pain. *J Electromyogr Kinesiol* 2008;18:559-67.
5. Whittaker JL, Teyhen DS, Elliott JM, et al. Rehabilitative ultrasound imaging: understanding the technology and its applications. *J Orthop Sports Phys Ther* 2007;37:434-49.
6. Teyhen DS. Rehabilitative Ultrasound Imaging Symposium. *J Orthop Sports Phys Ther* 2006;36:A1-A17.
7. Ferreira PH, Ferreira ML, Hodges PW. Changes in recruitment of the abdominal muscles in people with low back pain - Ultrasound measurement of muscle activity. *Spine* 2004;29:2560-6.
8. McMeeken JM, Beith ID, Newham DJ, et al. The relationship between EMG and change in thickness of transversus abdominis. *Clin Biomech* 2004;19:337-42.
9. Hodges PW, Pengel LHM, Herbert RD, et al. Measurement of muscle contraction with ultrasound imaging. *Muscle & Nerve* 2003;27:682-92.
10. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Aust J Physiother* 2006;52:91-102.
11. Terwee CB, Bot SDM, Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
12. de Vet HCW, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033-9.

13. Krebs DE. Declare your ICC type. *Phys Ther* 1986;66:1431.
14. Critchley DJ, Coutts FJ. Abdominal muscle function in chronic low back pain patients: measurement with real-time ultrasound scanning. *Physiotherapy* 2002;88:322-32.
15. Kidd AW, Magee S, Richardson CA. Reliability of real-time ultrasound for the assessment of the transversus abdominis function. *J Grav Physiol* 2002;9:P 131-2.
16. Mannion AF, Pulkovski N, Gubler D, et al. Muscle thickness changes during abdominal hollowing: an assessment of between-day measurement error in controls and patients with chronic low back pain. *Eur Spine J* 2008;17:494-501.
17. Norasteh A, Ebrahimi E, Salavati M, et al. Reliability of B-mode ultrasonography for abdominal muscles in asymptomatic and patients with acute low back pain. *J Body Movem Ther* 2007;11:17-20.
18. Teyhen DS, Gill NW, Whittaker JL, et al. Rehabilitative ultrasound imaging of the abdominal muscles. *J Orthop Sports Phys Ther* 2007;37:450-66.
19. Maher CG, Sherrington C, Herbert RD, et al. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther* 2003;83:713-21.
20. Toma V, Pulkovski N, Sprott H, et al. Reliability of measures of abdominal muscle thickness during abdominal hollowing as assessed with M-mode ultrasound. *Eur J Pain* 2006;10 S109.
21. Pietrek M, Sheikhzadeh A, Hagins M, et al. Evaluation of abdominal muscles by ultrasound imaging: reliability, and comparison to electromyography. *Eur Spine J* 2000;9:309.
22. Ferreira PH, Ferreira ML, Maher CG, et al. Clinical ultrasound test for transversus abdominus thickness: investigation of reliability 13th Biennial Conference - Musculoskeletal Physiotherapy Australia. Sydney, Australia, 2003.
23. Beazell JR, Grindstaff TL, Magrum EM, et al. Comparison of clinical test and real time ultrasound evaluation of muscle contraction in normals and patients with low back pain. *J Manual Manipulative Ther* 2006;14:168-9.
24. Bunce SM, Moore AP, Hough AD. M-mode ultrasound: a reliable measure of transversus abdominis thickness? *Clin Biomech* 2002;17:315-7.

25. Kiesel KB, Underwood FB, Mattacola CG, et al. A comparison of select trunk muscle thickness change between subjects with low back pain classified in treatment-based classification system and asymptomatic controls. *J Orthop Sports Phys Ther* 2007;37:596-607.
26. Rankin G, Stokes M, Newham DJ. Abdominal muscle size and symmetry in normal subjects. *Muscle Nerve* 2006;34:320-6.
27. Hides JA, Miokovic T, Belavy DL, et al. Ultrasound imaging assessment of abdominal muscle function during drawing-in of the abdominal wall: an intrarater reliability study. *J Orthop Sports Phys Ther* 2007;37:480-6.
28. Teyhen DS, Miltenberger CE, Deiters HM, et al. The use of ultrasound imaging of the abdominal drawing-in maneuver in subjects with low back pain. *J Orthop Sports Phys Ther* 2005;35:346-55.
29. Kiesel KB, Uhl T, Underwood FB, et al. Rehabilitative ultrasound measurement of select trunk muscle activation during induced pain. *Man Ther* 2008;13:132-8.
30. Teyhen DS, Rieger JL, Westrick RB, et al. Changes in deep abdominal muscle thickness during common trunk-strengthening exercises using ultrasound imaging. *J Orthop Sports Phys Ther* 2008;38:596-605.
31. John EK, Beith ID. Can activity within the external abdominal oblique be measured using real-time ultrasound imaging? *Clin Biomech* 2007;22:972-9.
32. Roddey TS, Brizzolara KJ, Cook KF. A comparison of two methods of assessing transverse abdominal muscle thickness in participants using real-time ultrasound in a clinical setting. *Orthop Phys Ther Pract* 2007;19:198-201.
33. Ainscough-Potts AM, Morrisey MC, Critchley DJ. The response of the transverse abdominis and internal oblique to different postures. *Man Ther* 2006;11:54-60.
34. Misuri G, Colagrande S, Gorini M, et al. In vivo ultrasound assessment of respiratory function of abdominal muscles in normal subjects. *Eur Respir J* 1997;10:2861-7.
35. Springer BA, Mielcarek BJ, Nesfield TK, et al. Relationships among lateral abdominal muscles, gender, body mass index, and hand dominance. *J Orthop Sports Phys Ther* 2006;36:289-97.

36. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
37. Chiradejnant A, Maher CG, Latimer J. Objective manual assessment of lumbar PA stiffness is now possible. *J Manipulative Physiol Ther* 2003;26:34-9.
38. Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. *Phys Ther* 1994;74:801-9.
39. Maher CG, Latimer J, Adams R. An investigation of the reliability and validity of posteroanterior spinal stiffness judgments made using a reference-based protocol. *Phys Ther* 1998;78:829-37.
40. Fleiss J. *The design and analysis of clinical experiments*. New York: Wiley, 1986.
41. Armstrong GD. The intraclass correlation as a measure of interrater reliability of subjective judgments. *Nurs Res* 1981;30:314-5.
42. Laschinger HKS. Intraclass correlations as estimates of interrater reliability in nursing research. *West J Nurs Res* 1992;14:246-9.

Tables and figures

Table 1. Search strategy

-
-
1. ultrasound OR ultra-sound OR ultra sound OR scanning OR imaging
 2. reliability OR repeatability OR test-retest OR assessment OR evaluation OR examination OR thickness OR activ\$ OR function OR change\$ OR investigation OR ICC OR limits of agreement OR critical difference
 3. Transversus abdominis OR internal oblique OR external oblique OR abdominal muscle\$
 4. 1 and 2
 5. 3 and 4
-

Table 2. Data extraction

Description of the sample
Sample size
Ultrasound mode (B or M mode)
Task performed by the participants (e.g. abdominal hollowing, rest)
Muscles investigated (e.g. Transversus abdominis, internal oblique, external oblique)
Length of the interval between the RUSI assessments
Blinding
Ordering of the tests (e.g. alternation, randomization, etc)
Description of the assessor
Description of the type of reliability (i.e. intra/inter rater, intra/interimage)
Reliability and agreement values (i.e. Intraclass Correlation Coefficient (ICC)*, Kappa, Standard Error of the Measurement -SEM, Coefficient of Variation -CV, Bland and Altman Plots, etc)

* Ten authors from studies that did not specify the type of ICC were contacted by email and we received responses from 6 of them^{14,15,22,23,26,32}.

Table 3. Item 6 (reproducibility) of the *Quality criteria for measurement properties* ¹¹

<i>Property</i>	<i>Definition</i>	<i>Quality criteria</i>
Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error).	+ ICC or Kappa ≥ 0.70 [¶] ;
		? doubtful design or method* (e.g. time interval not mentioned, inadequate description of the ICC tests);
		- ICC or Kappa < 0.70 , despite adequate design and method;
		0 no information found on reliability.
Agreement**	The extent to which the scores on repeated measures are close to each other (absolute measurement error).	+ MIC $>$ SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable;
		? doubtful design or method* OR (MIC not defined AND no convincing arguments that agreement is acceptable);
		- MIC \leq SDC OR MIC inside LOA, despite adequate design and method;
		0 no information found on agreement.

+ = positive rating; ? = indeterminate rating; - = negative rating; 0 = no information available

*Doubtful design or method = lacking a clear description of the design or methods of the study, sample size smaller than 50 subjects (should be at least 50 in every (subgroup) analysis)³⁶, or any important methodological weakness in the design or execution of the study.

[¶] In case of multiple reliability tests, the study will be rated + only if 75% or more of tests achieved the benchmark of 0.70.

ICC = Intraclass correlation coefficient; MIC = minimal important change; SDC = smallest detectable change; LOA = limits of agreement.

** MIC was not considered in the ratings of agreement as MIC is more related to self-report patient outcome measures than physiological measures such as RUSI.

Table 4. Description of the eligible studies

Study	Are patients seeking care for LBP?	Description of the sample	Description of the assessor	Ultra sound mode	Muscle task	Muscles investigated	Interval	Order of tests	blinding
Aiscough-Potts et al. 2006 ³³	No	30 healthy subjects (Physiotherapy students and staff with no history of LBP in the past 6 months) but only 10 subjects were required for the reliability testing.	No info	B	Automatic 1) Supine, 2) sitting on a chair, 3) sitting on a ball, 4) sitting on a ball while lifting the left foot by approximately 10 cm. The images were taken at the end of inspiration and expiration.	TrA, OI (right side)	No info	Random order using a random square Latin table	No info
Bezell et al. 2006 ²³	Mixed	19 healthy and 20 patients with LBP.	No info	No info	Voluntary 1) Abdominal drawing-in manoeuvre, 2) "...first four components of the Abdominal Muscle Strength Test (AMST)."	TrA, OI, OE (right side for healthy subjects and affected side for LBP patients)	"over three treatment sessions	No info	No info
Bunce et al. 2002 ²⁴	No	22 healthy adults with no history of LBP in the past 6 months.	No info	M	Automatic 1) Supine, 2) Standing, 3) Walking in a 3kph treadmill	TrA	"three separate days"	No info	No info
Critchley & Coultts 2002 ¹⁴	No	"10 different subjects."	No info	B	Rest images only (10 times/patient) in four-point kneeling.	TrA, OI, OE	Immediately after	No info	No info
Ferreira et al. 2003 ²²	Yes	20 low back pain patients.	"One of the testers was trained for 3 months prior to data collection while the other had no previous training"	B	No info	TrA	No info	No info	No info

Continue over the next page

Study	Are patients seeking care for LBP?	Description of the sample	Description of the assessor	Ultra sound mode	Muscle task	Muscles investigated	Interval	Order of tests	blinding
Hides et al. 2007 ²⁷	No	19 healthy subjects with no previous history of LBP.	“A physiotherapist underwent training and performed all of subsequent measurements”	B	Voluntary “Subjects were positioned in a supine hook-lying position with their hips in 45 degrees of flexion.” 1) Rest, 2) Abdominal draw in manouver (the procedure was repeated 6 times/patient).	TrA, OI (both sides)	1) Measured the same image 3 times (i.e. no interval) 2). Comparison of 3 images from the same task, same day (i.e. interval = immediately after). 3) 4-7 days interval.	Randomized	No info
John & Beith 2007 ³¹	No	24 healthy subjects.	No info	B and M	Automatic. Rest in crook lying position.	OE	Immediately after (3 images)	No info	No info
Kidd et al. 2002 ¹⁵	No	11 subjects without history of LBP.	No info	M	Abdominal draw in manouver (sitting and standing – 4 times each task).	TrA	“ 2 separate occasions”	No info	No info
Kiesel et al. 2008 ²⁹	No	8 asymptomatic subjects.	No info	B	Voluntary. Abdominal wall draw-in manouver in prone position (rest and activated).	TrA	No info	No	No info
Kiesel et al. 2007 ²⁵	56 patients were seeking care + 20 controls	15 subjects from the main study (unclear if they were controls or cases).	No info	B	Abdominal draw in manouver (rest and activated) in supine hook lying position.	TrA	Same day	No info	Yes

Continue over the next page

Study	Are patients seeking care for LBP?	Description of the sample	Description of the assessor	Ultra sound mode	Muscle task	Muscles investigated	Interval	Order of tests	blinding
Mannion et al. 2008 ¹⁶	14 seeking care + 14 controls	14 patients with chronic LBP + 14 controls.	No info	M	Voluntary. Abdominal hollowing in supine hook lying position (hips in 30 degrees of flexion) (rest and activated).	TrA, OI, OE	7 ± 2 days	No info	No info
McMeeken et al. 2004 ⁸	No	13 healthy subjects.	No info	B and M	Voluntary. Abdominal hollowing in supine lying and knees bent to 20 degrees of flexion (8 times).	TrA	7 days	Randomized	No info
Misuri et al. 1997 ³⁴	No	6 healthy male subjects.	No info	B	Voluntary. Seated in a high-backed armchair at 90 degrees. 1) Functional residual capacity; 2) Breath holding at residual volume; 3) Total lung capacity.	TrA, OI, OE, RA	No info	No info	No info
Norasteh et al. 2007 ¹⁷	No	27 healthy subjects + 12 Acute low back pain patients.	No info	B	Automatic. 1) Supine; 2) Standing and ; 3) Sitting.	TrA, OI, OE, RA	7 days	No info	No info
Pietrek et al. 2000 ²¹	No	12 healthy subjects.	No info	No info	Voluntary. Isometric trunk flexion, extension and rotation tasks with 4 levels of exertion.	TrA, OI, OE	No info	Randomized	No info
Rankin et al. 2006 ²⁶	No	10 subjects.	No info	B	Rest.	TrA, OE, OI	7 days	No info	No info
Roddey et al. 2007 ³²	No	70 healthy subjects.	No info	B	Voluntary. Lying supine with knees flexed. Rest and contracting TrA by "...performing pelvic floor contraction or bracing contraction.	TrA	Immediately after	No info	No info

Continue over the next page

Study	Are patients seeking care for LBP?	Description of the sample	Description of the assessor	Ultrasound mode	Muscle task	Muscles investigated	Interval	Order of tests	blinding
Springer et al. 2006 ³⁵	No	32 healthy department of defence beneficiaries (no history of LBP in the last 3 years).	“The examiner consisted of a senior physical therapist with 18 years of clinical experience and a student, who were provided a training session”	B	Voluntary. Abdominal draw in manoeuvre and rest in supine hook-lying posture.	TrA and total abdominal thickness.	Immediately after.	Randomized	Not for the measurements (The assessors were blinded to the treatment allocation).
Teyhen et al. 2005 ²⁸	Yes	30 subjects who had been seeking care for LBP within the previous 3 months.	“...physical therapy students who were provided a 3-hour training session in the measurement procedures”	B	Rest in supine hook-lying position. The reliability was calculated with TrA images in rest + thickness of total abdominal muscle.	TrA, and total lateral abdominal muscle thickness (TrA+OI+OE).	Immediately after	No info	No blinding for images.
Teyhen et al. 2008 ³⁰	No	10 normal subjects	No info	B	Abdominal draw in manoeuvre; abdominal crunch; abdominal sit-back; quadruped opposite upper and lower extremity lift.	TrA and OI	Same day	No info	No info
Toma et al. 2006 ²⁰	No	16 subjects	No info	M	Rest and abdominal hollowing in lying position (5 times). Also total muscle thickness at max TrA contraction.	TrA, OI and OE	1 day interval	No Info	No info

Notes: TrA transversus abdominis; OI internal oblique; OE external oblique; M motion mode; B bright mode; M motion mode; LBP low back pain

Table 5. Reproducibility values for each study

Study	Reliability for single measures	Reliability for changes in thickness	Notes
Aisncough-Potts et al. 2006 ³³	Intrater reliability (ICC) Inspiration: 0.97 Expiration: 0.99	No info	The study did not specify which muscles the ICC values pertain to. ICC type not specified.
Beazell et al. 2006 ²³	Intrater reliability (ICC _{3,1}) 0.94-0.99	No info	The study did not specify which muscles the ICC values pertain to.
Bunce et al. 2002 ²⁴	Intrater reliability TrA (ICC _{1,1} ; SEM (mm)) Supine: 0.94; 0.35 Standing: 0.88; 0.66 Walking: 0.88; 0.56	TrA ICC _{1,1} between supine and standing: 0.78 between supine and walking: 0.48	
Critchley & Coutts 2002 ¹⁴	Intrater reliability (ICC _{1,1} ; SEM (mm)) OE: 0.95; 0.66 OI: 0.98; 0.80 TrA: 0.94; 0.60	No info	
Ferreira et al. 2003 ²²	Inter-rater reliability Same image (ICC _{2,1}) 0.70-0.98	Intrater reliability (ICC _{2,1}) Experienced rater: 0.85 Inexperienced rater: 0.28 Inter rater reliability: 0.26	The study did not specify which muscles the ICC values pertain to.
Hides et al. 2007 ²⁷	Intra-rater reliability - Same image (ICC _{3,1} 95%CI ; SEM(mm)) IO rest: 0.99 (0.97-1.00); 0.028 IO contracted: 0.98 (0.95-1.00); 0.033 TrA rest: 0.98 (0.95-1.00); 0.01 TrA contracted: 0.97 (0.98-1.00); 0.020 “across 3 images” (ICC _{3,4} 95%CI ; SEM(mm)) IO rest: 0.82 (0.55-0.95); 0.157 IO contracted: 0.66 (0.23-0.92); 0.385 TrA rest: 0.62 (0.32-0.85); 0.247 TrA contracted: 0.80 (0.56-0.93); 0.224 2 days interval (ICC _{3,6} 95%CI ; SEM(mm)) IO rest: 0.69 (0.30-0.92); 0.295 IO contracted: 0.63 (0.21-0.94); 0.432 TrA rest: 0.85 (0.42-0.98); 0.089 TrA contracted: 0.84 (0.52-0.96); 0.176	No info	
John & Beith 2007 ³¹	OE Intrater reliability (ICC): 0.92	No info	ICC type not specified.

Study	Reliability for single measures	Reliability for changes in thickness	Notes	
Kidd et al. 2002 ¹⁵	<p><u>TrA same day</u> (ICC_{1,1}; SEM (mm); SEM (%)) “ range from 0.79 to 0.96” ; 0.29 to 0.57; 3.7% - 8.9%</p> <p><u>TrA Between days</u> (ICC_{1,1}; SEM (mm); SEM (%)) Sitting: 0.96; 0.18; 1.2 Standing: 0.88; 0.33;3.6</p>	No info		
Kiesel et al. 2008 ²⁹	<u>TrA</u> ICC _{3,1} : 0.95	No info		
Kiesel et al. 2007 ²⁵	<p><u>TrA</u> (ICC_{3,3}(95%CI); SEM(cm); MDC(cm)) Rest: 0.98 (0.91-0.99); 0.01; 0.03 Contracted: 0.97 (0.91-0.98); 0.02; 0.06</p>	<p><u>TrA % change</u>: ICC_{3,3}(95%CI); SEM(%); MDC(%) 0.96 (0.91-0.99); 6.26; 17.34</p>		
Mannion et al. 2008 ¹⁶	<p><u>Controls</u> (ICC_{3,1}; SEM(mm); CV) TrA rest Left: 0.86; 0.40; 10.7 Right: 0.83; 0.40; 10.3 TrA Max Left: 0.75; 0.65; 12.0 Right: 0.78; 0.58; 10.7 OI rest Left: 0.94; 0.72; 9.8 Right: 0.92; 0.58; 8.8 OE rest Left: 0.59; 1.03; 19.6 Right: 0.26; 0.84; 17.0</p>	<p><u>Low back pain patients</u> (ICC_{3,1}; SEM(mm); CV) TrA rest Left: 0.63; 0.46; 11.5 Right: 0.89; 0.27; 7.2 TrA Max Left: 0.41; 0.78; 14.3 Right: 0.88; 0.41; 7.7 OI rest Left: 0.85; 0.68; 9.8 Right: 0.73; 0.82; 10.8 OE rest Left: 0.51; 0.84; 14.6 Right: 0.42; 1.20; 19.1</p>	<p><u>Low back pain patients</u> (ICC_{3,1}; SEM(mm);CV) TrA contraction ratio Left: 0.28; 0.16; 11.6 Right: 0.80; 0.09; 6.0 OE contraction ratio Left: 0.57; 0.05; 5.6 Right: 0.43; 0.05; 5.4 OE+OI Contraction ratio Left: 0.39; 0.05; 5.6 Right: 0.25; 0.05; 4.5 TrA preferential activation ratio Left: 0.32; 0.03; 49.5 Right: 0.48; 0.02; 27.4</p>	
McMeeken et al. 2004 ⁸	<p><u>TrA</u> (ICC; 95% CI) B mode: 0.99 (0.96 - 0.99) M mode: 0.98 (0.94 - 0.99) Linear versus curvilinear transducers: 0.82 (0.87 - 0.96) Bland and Altman tests (mean difference (mm); SE of difference (mm); 95% limits of agreement (mm); repeatability coefficient. B mode: 0.03; 0.03; -0.17 - 0.24; 0.023 M mode: 0.04; 0.04; -0.23 - 0.03; 0.038 Linear versus curvilinear transducers -0.14; 0.14; -0.95 - 0.66; 0.045</p>	No info	ICC type not specified.	

Study	Reliability for single measures	Reliability for changes in thickness	Notes
Misuri et al. 1997 ³⁴	CV ranging from 0 to 15.7% “...for each muscle, between subject variability was highly significant (F ranging 9.1-273, p-values ranging from 0.003-0.00001) whereas intrasubject variability was not (F ranging 0.54-3.2, p-values ranging from 0.66-0.06).”	No info	The study presents CV for individual patients for each task.
Norasteh et al. 2007 ¹⁷	Asymptomatic subjects <u>ICC_{1,2}</u> ; <u>ICC_{1,3}</u> ; SEM OE: 0.96; 0.72; 0.33 OI: 0.97; 0.91; 0.07 TrA: 0.81; 0.80; 0.45 RA: n/a; 0.85; 0.84 Acute low back pain patients <u>ICC_{2,1}</u> + SEM OE: 0.87; 0.35 OI: 0.87; 0.31 TrA: 0.91; 0.30	No info	No unit of measurement was provided.
Pietrek et al. 2000 ²¹	Pearson's r: 0.55 - 0.97	No info	
Rankin et al. 2006 ²⁶	Between scans <u>ICC_{1,1}</u> 0.98-0.99 (95% CI 0.91-1.00) Between days <u>ICC_{1,2}</u> 0.96-0.99 (95% CI 0.85-1.00) “Bland and Altman tests produced mean differences close to zero and SD diff values were very low” 95% limits of agreement (cm) OI: 0.22 OE: 0.12 TrA: 0.09	No info	
Roddey et al. 2007 ³²	Deferred Assessment of TrA <u>ICC_{2,1}</u> ; SEM (mm) Right relaxed: 0.84; 0.04 Right contracted: 0.83; 0.07 Left relaxed: 0.90; 0.03 Left contracted: 0.91; 0.06 <u>Immediate Assessment of TrA</u> <u>ICC_{2,1}</u> ; SEM (mm) Right relaxed: 0.83; 0.03 Right contracted: 0.81; 0.09 Left relaxed: 0.93; 0.02 Left contracted: 0.92; 0.04	No info	

Continue over the next page

Study	Reliability for single measures	Reliability for changes in thickness	Notes
Springer et al. 2006 ³⁵	<p>TrA single measure ICC_{2,1}; 95% CI; SEM (mm) Rest: 0.93 (0.86-0.96); 0.32 Contracted: 0.96 (0.92-0.98) 0.45 TrA average of 3 measures ICC_{2,3}; 95% CI; SEM (mm): Rest: 0.98 (0.92-0.99); 0.13 Contracted: 0.99 (0.98-0.99); 0.20 TrA+OE+OI single measure ICC_{2,1}; 95% CI; SEM (mm) Rest: 0.98 (0.96-0.99); 0.80 Contracted: 0.99 (0.98-1.00); 0.71 TrA+OE+OI average of 3 measures ICC_{2,3}; 95% CI; SEM (mm) Rest: 1.00 (0.99-1.00); 0.35 Contracted: 1.00 (0.99-1.00); 0.34</p>	<p>TrA/total single measure (ICC_{2,1};95% CI ; SEM (%)) Rest: 0.91 (0.82-0.95); 2 Contracted: 0.98 (0.96-0.99); 1.2 TrA/total average of 3 measures (ICC_{2,3};95% CI ; SEM (%)) Rest: 0.99 (0.97-0.99); 0.5 Contracted: 0.99 (0.99-1.00); 0.7</p>	<p>The ratio presented (TrA/total) were calculated from values of the same image.</p>
Teyhen et al. 2005 ²⁸	<p>TrA ICC_{3,1}; 95% CI; SEM(cm); CV Intra-image: 0.98 (0.96-0.99); 0.013; 5 Inter-image: 0.93 (0.77-0.99); 0.031; 11 TrA+OE+OI ICC_{3,1}; 95% CI; SEM(cm); CV Intra-image: 0.99 (0.99-1.0); 0.018; 5 Inter-image: 0.97 (0.77-0.99); 0.087; 14 ICC_{2,2} "greater or equal to 0.95" SEM (mm) TrA: 0.09 OI: 0.29</p>	<p>No info</p>	<p>The values from TrA and total muscle thickness (TrA+OE+OI thickness) were obtained from the same image</p>
Teyhen et al. 2008 ³⁰	<p>ICC_{2,2} "greater or equal to 0.95" SEM (mm) TrA: 0.09 OI: 0.29</p>	<p>No info</p>	<p>The values from TrA and total muscle thickness (TrA+OE+OI thickness) were obtained from the same image</p>
Toma et al. 2006 ²⁰	<p>TrA (ICC; SEM (mm);SEM (%)) 0.70-0.94; 0.44-0.74; 8-15</p>	<p>TrA ICC Left: 0.44 Right: 0.70</p>	<p>ICC type not specified</p>

Notes: TrA transversus abdominis; OI internal oblique; OE external oblique; IC internal oblique; OE external oblique; B bright mode; M motion mode; LBP low back pain
ICC intraclass correlation coefficient; SEM standard error of the Measurement; CV coefficient of variation;

Table 6. Quality of reliability (for thickness and thickness changes) and agreement rated by the Quality Criteria

Reference	Reliability thickness	Reliability changes in thickness	Agreement	notes
Aisncough-Potts et al. 2006 ³³	?	0	0	No time interval provided / did not provide ICC type / underpowered.
Beazell et al. 2006 ²³	?	0	0	Underpowered.
Bunce et al. 2002 ²⁴	?	?	?	Only 50% of the ICC values were above 0.70 / underpowered.
Critchley & Coutts 2002 ¹⁴	?	0	?	Underpowered.
Ferreira et al. 2003 ²²	?	-	0	No time interval provided / only 33% of the ICC values were above 0.70 / underpowered.
Hides et al. 2007 ²⁷	-	0	?	Across 3 images and 2 days: only 50% of the ICC values were above 0.70 / underpowered.
John & Beith 2007 ³¹	?	0	0	Did not provide ICC type / underpowered.
Kidd et al. 2002 ¹⁵	?	0	?	Underpowered.
Kiesel et al. 2008 ²⁹	?	0	0	No time interval provided / underpowered.
Kiesel et al. 2007 ²⁵	?	?	?	Underpowered
Mannion et al. 2008 ¹⁶	?	-	0	ICC values for reliability of thickness changes were below 0.70 / underpowered.
McMeeken et al. 2004 ⁸	?	0	?	Did not provide ICC type / underpowered.
Misuri et al. 1997 ³⁴	?	0	0	Used CV only/ underpowered.
Norasteh et al. 2007 ¹⁷	?	0	?	Underpowered.
Pietrek et al. 2000 ²¹	?	0	0	Used Pearson's r / no time interval provided / underpowered.
Rankin et al. 2006 ²⁶	?	0	?	Underpowered.
Roddey et al. 2007 ³²	?	0	?	The time interval (immediately after) could inflate the reliability.
Springer et al. 2006 ³⁵	?	?	?	The time interval (immediately after) could inflate the reliability /underpowered.
Teyhen et al. 2005 ²⁸	?	0	?	The time interval (immediately after) could inflate the reliability / underpowered.
Teyhen et al. 2008 ³⁰	?	0	?	Underpowered.
Toma et al. 2006 ²⁰	?	?	?	Did not provide ICC type/ underpowered.

Notes: ICC intraclass correlation coefficient; CV coefficient of variation.

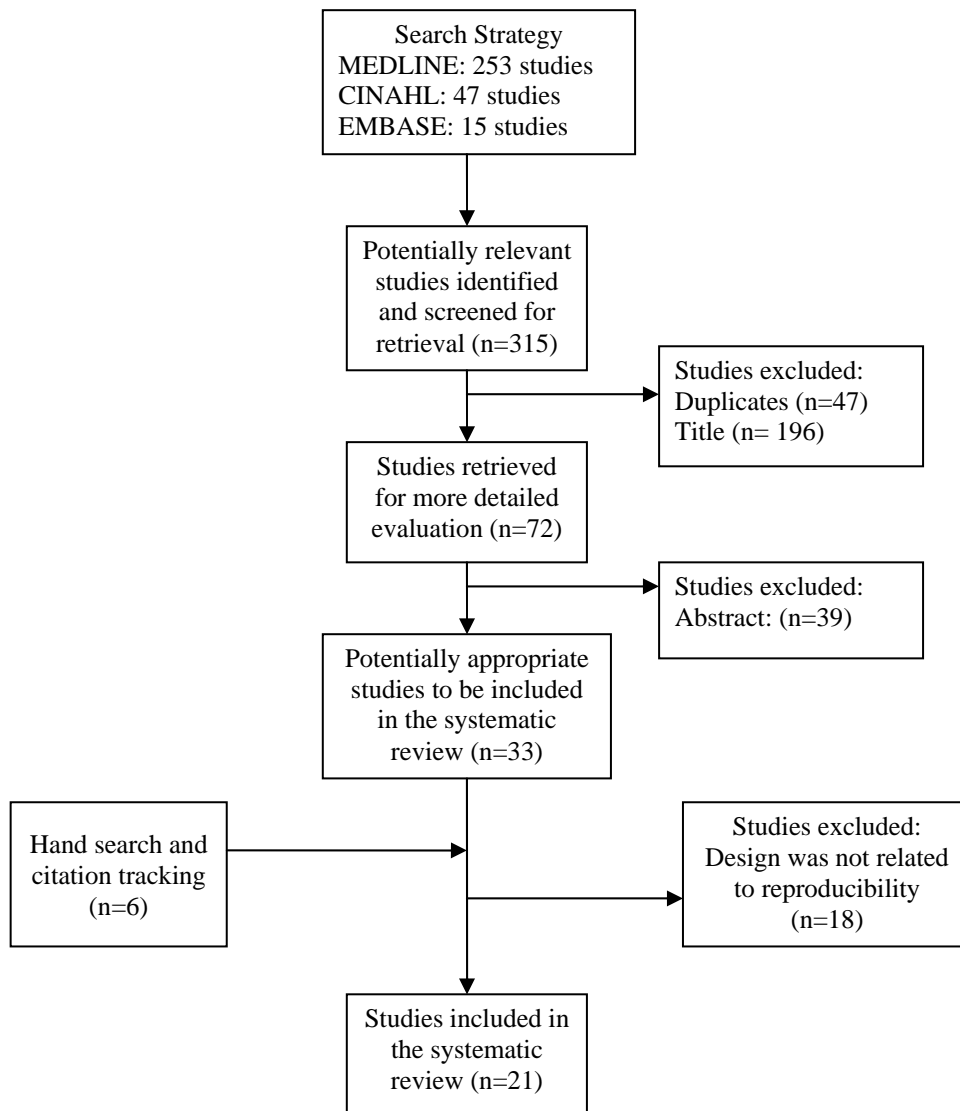


Figure 1. QUORUM (Quality of Reporting of Meta-analyses) statement flow diagram of the literature search.

Chapter Five

An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain

Chapter Five is published as:

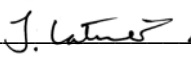
Costa LOP, Maher CG, Latimer J, Hodges PW, Shirley D (2009) An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain. Submitted for publication to *European Spine Journal* on 08.01.09.

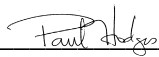
**Statement from co-authors confirming authorship contribution of the
PhD candidate**

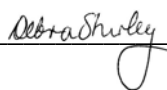
As co-authors of the paper “An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain”, we confirm that Leonardo Costa has made the following contributions:

- Conception and design of the research
- Data collection
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

Christopher G Maher _____  Date 20.03.2009

Jane Latimer _____  Date 20.03.2009

Paul W Hodges _____  Date 20.03.2009

Debra Shirley _____  Date 20.03.2009

An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain

Leonardo O P Costa^{1*}, Chris G Maher¹, Jane Latimer¹, Paul W Hodges², Debra Shirley³

¹ Musculoskeletal Division, The George Institute For International Health and Faculty of Medicine, The University of Sydney, Australia

² NHMRC Centre of Clinical Research Excellence in Spinal Pain, Injury and Health, School of Health of Rehabilitation Sciences, The University of Queensland, Australia

³ Faculty of Health Sciences, The University of Sydney, Australia

* Corresponding author: Leonardo Oliveira Pena Costa, The George Institute For International Health, PO Box M201 Missenden Road NSW 2050 Australia, licos3060@gmail.com, phone: 612 82382440; Fax 612 96570301.

Sources of support: This study was funded by the Physiotherapy Research Foundation, Australian Physiotherapy Association. Leonardo O P Costa is a PhD student supported by CAPES – Ministério da Educação – Brazil and Pontifícia Universidade Católica de Minas Gerais – Brazil; Chris G Maher and Paul W Hodges hold research fellowships funded by the National Health and Medical Research Council of Australia.

Abstract: Ultrasound (US) measures are used by clinicians and researchers to evaluate improvements in activity of the abdominal muscles in patients with low back pain. Studies evaluating the reproducibility of these US measures provide some information, however little is known about the reproducibility of these US measures over time in patients with low back pain. The objectives of this study were to estimate the reproducibility of ultrasound measurements of automatic activation of the lateral abdominal wall muscles using a leg force task in patients with chronic low back pain. Thirty-five participants from an existing clinical trial participated. A reproducibility analysis was undertaken from all patients using data collected at baseline and after treatment. The reproducibility of measurements of thickness, muscle activation (thickness changes) and muscle improvement/deterioration after intervention (differences in thickness changes from single images made before and after treatment) was analysed. The reproducibility of static images (*thickness*) was excellent ($ICC_{2,1} = 0.97$, 95% CI = 0.96-0.97, Standard Error of the Measurement (SEM) = 0.04cm, Smallest Detectable Change (SDC) = 0.11cm), the reproducibility of *thickness changes* was moderate ($ICC_{2,1} = 0.72$, 95% CI 0.65-0.76, SEM = 15%, SDC 41%), while the reproducibility of *differences in thickness changes* from single images with statistical adjustment for duplicate measures was poor ($ICC_{2,1} = 0.44$, 95% CI 0.33-0.58, SEM = 21%, SDC = 66.5%). Improvements in the testing protocol must be performed in order to enhance reproducibility of US as an outcome measure for abdominal muscle activation.

Key words: Ultrasound, abdominal muscles, reproducibility, agreement, reliability.

Introduction

The use of motor control exercise (also known as specific stabilisation exercise) in the treatment of low back pain has become widespread[10, 12, 21, 22]. The rationale for the use of motor control exercise is that the deep abdominal and paraspinal muscles have a critical role in dynamic control of the lumbar spine[17, 18]. For example, delayed onset of activity of the Transversus abdominis muscle (TrA) has been reported in patients with recurrent low back pain[16, 17] compared to asymptomatic subjects. Most studies that have measured the activity of the deep spinal muscles use fine-wire electromyography (EMG), which is costly, time-consuming, potentially uncomfortable, and includes risks such as infection. An alternative approach is to measure the recruitment of the muscles indirectly by assessment of morphologic changes of the muscles (i.e. thickness changes) using real-time ultrasound imaging (US)[15].

If ultrasound measures of abdominal muscle activation are to be useful they need to have acceptable reproducibility. Reproducibility is defined as the degree to which repeated measurements provide similar results and will be used in this manuscript as an umbrella term for reliability and agreement. Agreement assesses how close the results of repeated measurements are, by estimating the measurement error in repeated measurements. Reliability assesses whether study subjects could be distinguished from each other, despite measurement errors[8, 31].

The available studies that have evaluated the reproducibility of US measures of abdominal muscle activation answer some, but not all, of the questions about the use of this test to guide the clinical management of low back pain. The reproducibility of

US measures for abdominal wall muscles was extensively discussed in a recent systematic review[6]. Most of the 21 eligible studies recruited healthy, young participants, and only three studies recruited participants with low back pain[13, 26, 32]. These three studies provide good evidence that US can provide reproducible measures of *thickness* of the abdominal muscles, which was a consistent finding in the remaining studies of subjects without low back pain[2, 3]. Another important difference among the studies was the task performed by the participants during the US measures. In most of the available studies participants were asked to voluntarily contract the abdominal muscles (i.e. abdominal “draw-in” manouever)[26, 30, 32, 33]. Only a few studies used simple tasks to automatically activate the abdominal muscles (e.g. asking patients to move their legs or contract the leg muscles isometrically while the images were made from the abdominal wall)[4, 13].

Physiotherapists commonly use US for measurement of either the *activation* of the muscle (i.e. the difference between the thickness from an image with the muscle activated and the thickness from an image with the muscle at rest – referred to as *thickness changes*) or the *improvement in activation* of the muscle after intervention (i.e. as an outcome measure of improvement – referred to as *difference in thickness changes over time*). Only two previous studies have investigated the reproducibility of measures of *thickness changes* of the abdominal muscles in patients with low back pain[13, 26]. One study that tested reproducibility of US using a leg force task to produce automatic changes of the abdominal muscles, concluded that the *thickness changes* measures were only reproducible if the examiner was highly experienced[13]. The second study testing the reproducibility of US measures in patients performing the abdominal “draw in” manouever found a wide variety of

results ranging from poor to excellent reproducibility[26]. To date no study has investigated the reproducibility of the *difference in thickness changes over time*. Hence, it is still unclear if US measures are reproducible for the most important measures in a clinical population, that is, measurement of muscle activation using *thickness changes* and also measurement of improvement/deterioration of muscle activation using *differences in thickness changes over time*.

The objective of this study was to estimate the reproducibility of ultrasound measures of automatic activation of the lateral abdominal wall muscles during a leg force task in patients with chronic non-specific low back pain.

Methods

This study was nested within an existing clinical trial that compared the efficacy of motor control exercise (MCE) versus placebo in patients with chronic non-specific low back pain[25]. From the main study sample (n=154), a sub-sample of the last 35 participants was selected in order to test the automatic recruitment of the abdominal wall muscles by real-time ultrasound imaging. Eleven of the thirty-five patients refused to be tested after the intervention period leaving 24 patients for the post-intervention follow up (12 in the MCE group and 12 in the placebo group). The characteristics of the participants are presented in table 1. The study design, procedures and informed consent were approved by The University of Sydney Human Research Ethics Committee.

Eligibility Criteria

Participants were included if they had non-specific low back pain of at least 3 months duration, were currently seeking care for low back pain, were aged greater than 18 and less than 80 years, comprehended English, and they expected to continue residing within the study region for the study duration. Exclusion criteria were: suspected or confirmed serious spinal pathology, pregnancy, nerve root compromise, previous spinal surgery, scheduled for major surgery during treatment or follow up period, and presence of any contraindication to exercise[1]. We also excluded participants who were able to activate their TrA muscle for longer than 10 seconds (as preliminary evidence from previous trial suggested that these patients were less likely to benefit from a MCE program)[25].

Procedures

Ultrasound measurements

The procedures used in this study followed a previously published protocol[11]. Ultrasound images were made with a 10 cm, 5-10MHz linear wideband array transducer (Terason ultrasound systems, Teratech). The transducer was placed transversely across the abdominal wall at a point between the inferior angle of the rib cage and the iliac crest and ~10 cm from the umbilicus. This position was then adjusted by slightly moving the transducer head to ensure that the anteriomedial aspect of the TrA, including its medial edge, were visualised. Additionally, some gentle pressure was applied to the transducer head over the abdominal wall in order to ensure that the orientation of the muscle fibres was perpendicular to the transducer

head avoiding possible errors due to artefact anisotropy. The images were then frozen, saved and stored for later data extraction.

Participants were asked to perform a simple task which is expected to automatically activate the abdominal muscles and has been described in detail elsewhere[11].

Participants were positioned in supine with the hips flexed to $\sim 50^{\circ}$, and knees flexed to $\sim 90^{\circ}$ with the lower legs supported by slings around the knees and ankles.

Participants were instructed to perform isometric knee flexion followed by isometric knee extension. Two images were recorded from the left and then the right abdominal wall during each task; the first images taken with the muscles at rest and the second with the patient performing the isometric movement of the knee equivalent to 7.5% of body weight. Images were recorded at the end of the patient's expiration (patients were instructed to stop breathing without closing the glottis).

Two load cells were attached around the ankles in order to provide feedback to the patient about the target force. At each testing occasion 16 images were collected [2 tasks (4 images; being 2 at rest and 2 during activation) x 2 trials x 2 sides]. The order of the tasks and sides was counterbalanced.

Data extraction

The data from the images were extracted using custom-designed imaging software.

An electronic grid was placed over the image and the thickness measurements of three muscles (obliquus internus - OI, obliquus externus-OE and transversus abdominis-TrA) were made 1.0, 1.5 and 2.0 cm from the medial edge of the TrA (figure 1). The average of the three measures for each muscle was used for analysis; the thickness change was expressed as a proportion of thickness at rest.

We estimated the reproducibility of *thickness* of ultrasound measures by comparing the measures taken from the first and the second static image (i.e. we did not remove the patient from the plinth between the images). The reproducibility of *thickness changes* (reflecting the activation of the muscle) was calculated by comparing pairs of percentage changes in thickness during the activation tasks at baseline. Finally we calculated the reproducibility of *differences in thickness changes over time* (representing the improvement or deterioration in muscle activation) by comparing the differences in muscle activation of the first trial against the second trial (figure 2) (we also calculated the reproducibility of the *differences in thickness changes* across different combinations of baseline and post intervention scores, finding similar results). It is important to note that this study evaluated the reproducibility of single measures of *thickness change* and *differences in thickness change over time*. Because some studies average duplicate measures (to take into account some of the trial-to-trial variability) we estimated the reliability of such measures using the Spearman-Brown formula[9].

Statistical analysis

To describe reliability we calculated the Intraclass Correlation Coefficient ($ICC_{2,1}$) with 95% confidence intervals. We estimated the reliability of measures derived from the mean of two replicate measures using the Spearman-Brown formula. Our data fits the Spearman-Brown model assumptions (i.e. data was collected in a parallel (test-retest) design and the differences between the standard deviations of the 1st and the 2nd set of measures were less than 15%)[9]. A guideline for the use of the Spearman-Brown formula advocates that this adjustment technique is accurate for adjustments up to two replicate measures[9].

We used two measures of agreement: the Standard Error of the Measurement ($SEM_{consistency}$) and the Smallest Detectable Change (SDC). The SEM was calculated by dividing the standard deviation of the mean differences between the two measurements by $\sqrt{2}$ (i.e. $SEM = SD_{differences} / \sqrt{2}$). The SEM reflects the error of the instrument itself. The SDC was calculated by the formula $SDC = 1.96 \times \sqrt{2} \times SEM$. The SDC reflects the smallest within person change in a score that, with $P < 0.05$, can be interpreted as a “real” change, above measurement error in one individual.

Results

We found excellent reliability and agreement values for *thickness*, moderate reliability and agreement values for *thickness changes* and finally poor reproducibility for *differences in thickness changes over time*. The reliability coefficients (i.e. ICCs and adjusted ICCs) and agreement values (SEMs and SDCs) are presented in table 2.

Discussion

The use of ultrasound imaging by physiotherapists, as a feedback and measurement tool for patients with low back pain, has been increasing in the last decade. Our study aimed to estimate the reproducibility of ultrasound measurements of automatic activation of the lateral abdominal wall muscles during a leg force task in patients with chronic non-specific low back pain. The reproducibility scores were excellent for measuring *thickness*, moderate for *thickness changes* and poor for the *differences in thickness changes over time*.

In terms of reproducibility of measures of muscle thickness, the results of this study are consistent with the available literature on the topic, finding excellent ICC scores and also very small SEM and SDC scores[2-4, 7, 13, 14, 19, 20, 27-29, 32]. This demonstrates that the procedures used for obtaining US images of the abdominal wall and extracting data from the images are reproducible. Additionally it seems that the reliability and agreement of studies that recruited patients with low back pain are similar to the studies that recruited normal subjects. The reproducibility of *thickness changes* was lower compared to *thickness*, which is expected given the fact that two images (one with the muscle contracted and another with the muscle at rest) are required to determine the level of muscle activity (as reflected by *thickness change*). We also found a SEM of 15% and a SDC of 41.6% which means that the measurement error is around 15% and there would need to be 41.6% improvement in muscle activation to be sure that a true change had occurred. While measures of *thickness changes* taken from single images have only low reliability, we estimate that reliability would be moderate ($ICC_{2,1}=0.72$, 95% CI 0.65 to 0.77) for measures based upon the average of two measures. However measuring *differences in thickness changes over time* seems more difficult and even with statistical adjustment the reproducibility was poor ($ICC_{2,1}=0.44$, 95% CI 0.33 to 0.58).

Although we considered the agreement values (both SEM and SDC) for *thickness changes* and *differences in thickness changes over time* high, it is important to consider that there is no normative data in the literature that could allow us to determine what the minimum important change (MIC) is. The ideal scenario would be that the SDC is smaller than MIC[8], but unfortunately these values were not available in the literature.

To date, there is only one study that has estimated reliability of *thickness changes* for abdominal muscles in low back pain patients and asymptomatic controls[26]. The authors of this study measured the intra-rater reliability of ultrasound measurements on two separate days in patients performing “abdominal hollowing” (voluntary abdominal task which contrasts to our study in which the recruitment of the abdominal muscles was automatic). Voluntary tasks are more likely to have lower reproducibility as these tasks depend on skill and motivation of the participant, whereas automatic tasks do not have these potential confounders. Their data were analysed in many different ways, but the conclusions were very similar to our study: excellent reliability for *thickness* (ICC ranging from 0.75 to 0.94); but poor to moderate reliability for *thickness changes* (ICC ranging from 0.26 to 0.72). A potential limitation of our study was that the participants were not repositioned during the process of obtaining the images (i.e. we did not take the participants off the plinth) and therefore our reproducibility scores for *thickness* and *thickness changes* may be overestimated.

We attempted to improve the precision of the ultrasound measures firstly by training the examiner intensively, secondly by using a unique software that uses an electronic grid which avoids visual distortions (most of the previous studies used a “grid over the screen”[11], which is prone to distortions due to the angle of the screen); and finally by using an accurate load cell system in a stable frame which standardises the forces generated by participants performing the leg force task. Additionally we are confident that a clinically relevant population (i.e. patients seeking care for LBP) was selected to participate in the current study.

One potential limitation of this study was that we estimated the reproducibility for duplicate measures based upon single measures. We chose to collect single measures so that the reproducibility study could be accommodated within an existing clinical trial and because many clinicians use single ultrasound measures in clinical practice. While other studies have taken measures based upon the mean of up to 20 replicate measures[4] these estimates of reliability may be artificially high and not representative of clinical practice or be feasible for implementation. We believe that the mean of two duplicate measures is a feasible measurement protocol and used the Spearman-Brown formula to estimate the reliability for duplicate measures for operators who collect data in this way. Prior to using the adjustment we checked that the assumptions for the formula[9] held in our data set and so we are confident in the estimates of reliability we provide. Our data is the only data available for the reproducibility of the *differences in thickness changes over time*[6].

The whole process of performing US measures has multiple potential sources of error (e.g. accuracy of measurements of distance, identification of landmarks, ability to perform the tasks properly, and position of patient/transducer). Additionally it has to be acknowledged that trial to trial variation in performance of the activation tasks is expected. It would be useful to consider whether modifications in the test protocol may enhance the reproducibility of US measures (especially for *thickness changes* and *differences in thickness changes over time*). One approach that has been shown to enhance reliability of other low back assessments is to further standardise the protocol[5, 23, 24]. We feel the following aspects of the test may need further consideration in order to achieve better standardisation of our testing protocol:

- Although we performed the pre and post treatment measures at similar times of the day and asked the patients to empty their bladder before the US tests, there was no specific bowel or bladder preparation for the tests (i.e. diet restrictions, emptying the bowel etc). This could impact on the abdominal volumes, distending the abdominal muscles and therefore changing the thickness of the muscles during the tasks.
- Although we used a load-cell system that could identify whether the patient achieved the target force (i.e. 7.5% of the body weight), it may have been difficult for patients to repeat the tasks in a similar fashion to that performed at baseline. It is expected that the performance in tasks could vary between trials.
- Although previous studies found similar reliability values for Bright (B) and Motion (M) modes of US[19, 27], we think that the M mode could provide better reproducibility for *thickness changes*. The main reason for this is that the operator would acquire one image only when using the M mode, compared to acquiring two images when using the B mode.
- Increase the number of repeated measures, as this could decrease the variability, enhancing reproducibility. In this study only single trials were recorded and then adjusted statistically to estimate the benefit of multiple measures. Further studies are needed in order to establish whether an increased number of repetitions for this specific test can improve the reproducibility of the measure to a more acceptable level.

References

1. ACSM (1995) ACSM's Guidelines for exercise testing and prescription. Williams and Wilkins, Baltimore.
2. Ainscough-Potts AM, Morrisey MC, Critchley DJ (2006) The response of the transverse abdominis and internal oblique to different postures. *Man Ther* 11:54-60.
3. Beazell JR, Grindstaff TL, Magrum EM, Cullaty M, Hart JM, Shen FH (2006) Comparison of clinical test and real time ultrasound evaluation of muscle contraction in normals and patients with low back pain. *J Manual Manipulative Ther* 14:168-169.
4. Bunce SM, Moore AP, Hough AD (2002) M-mode ultrasound: a reliable measure of transversus abdominis thickness? *Clin Biomech* 17:315-317.
5. Chiradejnant A, Maher CG, Latimer J (2003) Objective manual assessment of lumbar PA stiffness is now possible. *J Manipulative Physiol Ther* 26:34-39.
6. Costa LOP, Maher CG, Latimer J, Smeets RJE (2009) Reproducibility of real-time ultrasound for the measurement of abdominal muscle activity: a systematic review. *Phys Ther* (accepted for publication).
7. Critchley DJ, Coutts FJ (2002) Abdominal muscle function in chronic low back pain patients: measurement with real-time ultrasound scanning. *Physiotherapy* 88:322-332.
8. de Vet HCW, Terwee CB, Knol DL, Bouter LM (2006) When to use agreement versus reliability measures. *J Clin Epidemiol* 59:1033-1039.
9. Feldt LS, Charter RA (2003) Estimating the reliability of a test split into two parts of equal or unequal length. *Psychol Methods* 8:102-109.
10. Ferreira ML, Ferreira PH, Latimer J, Herbert RD, Hodges PW, Jennings MD, Maher CG, Refshauge KM (2007) Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. *Pain* 131:31-37.
11. Ferreira PH, Ferreira ML, Hodges PW (2004) Changes in recruitment of the abdominal muscles in people with low back pain - Ultrasound measurement of muscle activity. *Spine* 29:2560-2566.
12. Ferreira PH, Ferreira ML, Maher CG, Herbert RD, Refshauge K (2006) Specific stabilisation exercise for spinal and pelvic pain: A systematic review. *Aust J Physiother* 52:79-88.
13. Ferreira PH, Ferreira ML, Maher CG, Refshauge K, Latimer J (2003) Clinical ultrasound test for transversus abdominis thickness: investigation of reliability In: 13th Biennial Conference - Musculoskeletal Physiotherapy Australia. Sydney, Australia.

14. Hides JA, Miokovic T, Belavy DL, Stanton WR, Richardson CA, Hides JA, Miokovic T, Belavy DL, Stanton WR, Richardson CA (2007) Ultrasound imaging assessment of abdominal muscle function during drawing-in of the abdominal wall: an intrarater reliability study. *J Orthop Sports Phys Ther* 37:480-486.
15. Hodges PW, Pengel LHM, Herbert RD, Gandevia SC (2003) Measurement of muscle contraction with ultrasound imaging. *Muscle Nerve* 27:682-692.
16. Hodges PW, Richardson CA (1999) Altered trunk muscle recruitment in people with low back pain with upper limb movement at different speeds. *Arch Phys Med Rehabil* 80:1005-1012.
17. Hodges PW, Richardson CA (1998) Delayed postural contraction of transversus abdominis in low back pain associated with movement of the lower limb. *J Spinal Disord* 11:46-56.
18. Hodges PW, Richardson CA (1996) Inefficient muscular stabilization of the lumbar spine associated with low back pain - A motor control evaluation of transversus abdominis. *Spine* 21:2640-2650.
19. John EK, Beith ID (2007) Can activity within the external abdominal oblique be measured using real-time ultrasound imaging? *Clin Biomech* 22:972-979.
20. Kidd AW, Magee S, Richardson CA (2002) Reliability of real-time ultrasound for the assessment of the transversus abdominis function. *J Grav Physiol* 9:131-132.
21. Macedo LG, Latimer J, Maher CG, Hodges PW, Nicholas M, Tonkin L, McAuley JH, Stafford R (2008) Motor control or graded activity exercises for chronic low back pain? A randomised controlled trial. *BMC Musculoskelet Disord* 9:65.
22. Macedo LG, Maher CG, Latimer J, McAuley J (2009) Motor control exercises for persistent nonspecific low back pain: A systematic review. *Phys Ther* 89:9-25.
23. Maher C, Adams R (1994) Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. *Phys Ther* 74:801-809.
24. Maher CG, Latimer J, Adams R (1998) An investigation of the reliability and validity of posteroanterior spinal stiffness judgments made using a reference-based protocol. *Phys Ther* 78:829-837.
25. Maher CG, Latimer J, Hodges PW, Refshauge KM, Moseley GL, Herbert RD, Costa LOP, McAuley J (2005) The effect of motor control exercises versus placebo in patients with chronic low back pain. *BMC Musculoskelet Disord* 6:54.
26. Mannion AF, Pulkovski N, Gubler D, Gorelick M, O'Riordan D, Loupas T, Schenk P, Gerber H, Spratt H (2008) Muscle thickness changes during abdominal hollowing: an assessment of between-day measurement error in controls and patients with chronic low back pain. *Eur Spine J* 17:494-501.

27. McMeeken JM, Beith ID, Newham DJ, Milligan P, Critchley DJ (2004) The relationship between EMG and change in thickness of transversus abdominis. *Clin Biomech* 19:337-342.
28. Norasteh A, Ebrahimi E, Salavati M, Rafiei J, Abbasnejad E (2007) Reliability of B-mode ultrasonography for abdominal muscles in asymptomatic and patients with acute low back pain. *J Body Movem Ther* 11:17-20.
29. Pietrek M, Sheikhzadeh A, Hagins M, Nordin M (2000) Evaluation of abdominal muscles by ultrasound imaging: reliability, and comparison to electromyography. *Eur Spine J* 9:309.
30. Springer BA, Mielcarek BJ, Nesfield TK, Teyhen DS (2006) Relationships among lateral abdominal muscles, gender, body mass index, and hand dominance. *J Orthop Sports Phys Ther* 36:289-297.
31. Terwee CB, Bot SDM, Boer MR, van der Windta DAWM, Knola DL, Dekkera J, Bouter LM, Veta HCW (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60:34-42.
32. Teyhen DS, Miltenberger CE, Deiters HM, Del Toro YM, Pulliam JN, Childs JD, Boyles RE, Flynn TW (2005) The use of ultrasound imaging of the abdominal drawing-in maneuver in subjects with low back pain. *J Orthop Sports Phys Ther* 35:346-355.
33. Teyhen DS, Rieger JL, Westrick RB, Miller AC, Molloy JM, Childs JD (2008) Changes in deep abdominal muscle thickness during common trunk-strengthening exercises using ultrasound imaging. *J Orthop Sports Phys Ther* 38:596-605.

Tables and figures

Table 1. Characteristics of study participants at baseline (n=35)

Variable	
Gender	
Female (%)	22 (62.9)
Age (years)	53.3 (11.27)
Low back pain duration (years)	6.2 (7.70)
Weight (kg)	69.3 (11.49)
Height (m)	1.6 (0.08)
Primary outcomes	
Pain Intensity Numerical Rating Scale (0-10)	6.6 (2.16)
Global Perceived Effect (-5 to + 5)	-2.2 (2.08)
Patient Specific Function Scale (0-30)	10.1 (5.84)
Secondary outcome	
Roland Morris Disability Questionnaire (0-24)	12.1 (5.45)

Continuous data are mean (SD), categorical data are N (%)

Table 2. Reliability and agreement statistics for ultrasound measures of *thickness*, *thickness changes* and *differences in thickness changes over time*.

	<i>Thickness</i>	<i>Thickness Changes</i>	<i>Differences in Thickness Changes Over Time</i>
N (Subjects)	35	35	24
N (Images)	840	420	288
Single measure ICC _{2,1} (95% CI)	not applicable	0.56 (0.49 to 0.62)	0.31 (0.20 to 0.41)
Average of 2 measures ICC _{2,1} (95% CI)	0.97 (0.96 to 0.97)	0.72 (0.65 to 0.77)*	0.44 (0.33 to 0.58)*
SEM	0.04 cm	15%	24%
SDC	0.11 cm	41.6%	66.5%

* adjusted by using the Spearman Brown formula[9].

ICC=Intraclass correlation coefficient; SEM=Standard error of the measurement;
SDC=Smallest detectable change

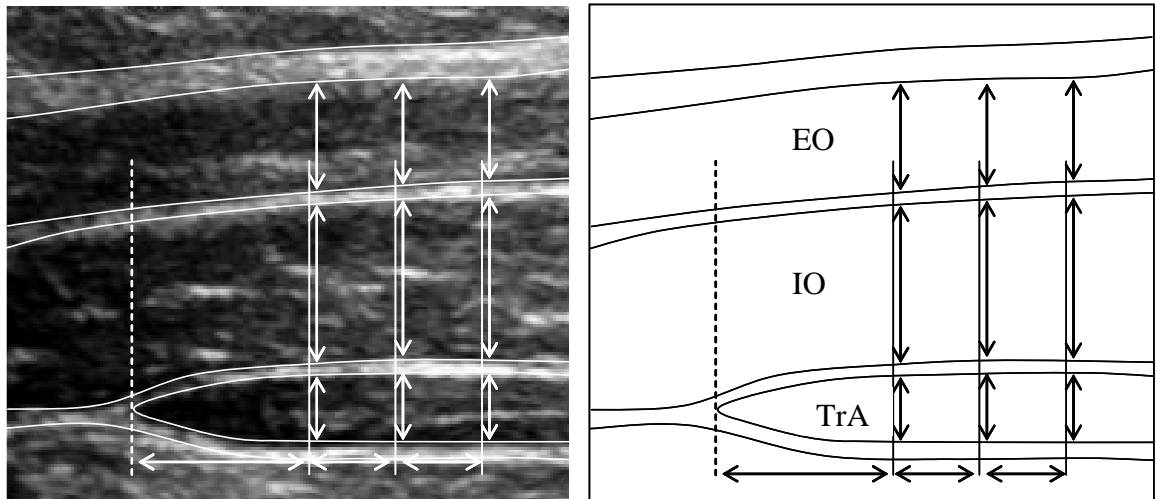


Figure 1. Ultrasound image of the abdominal muscles, the dashed line represents the edge of the Transversus abdominis muscle, the thickness measures are indicated by the vertical arrows.

EO= Obliquus externus; IO= Obliquus internus; TrA=Transversus Abdominis.

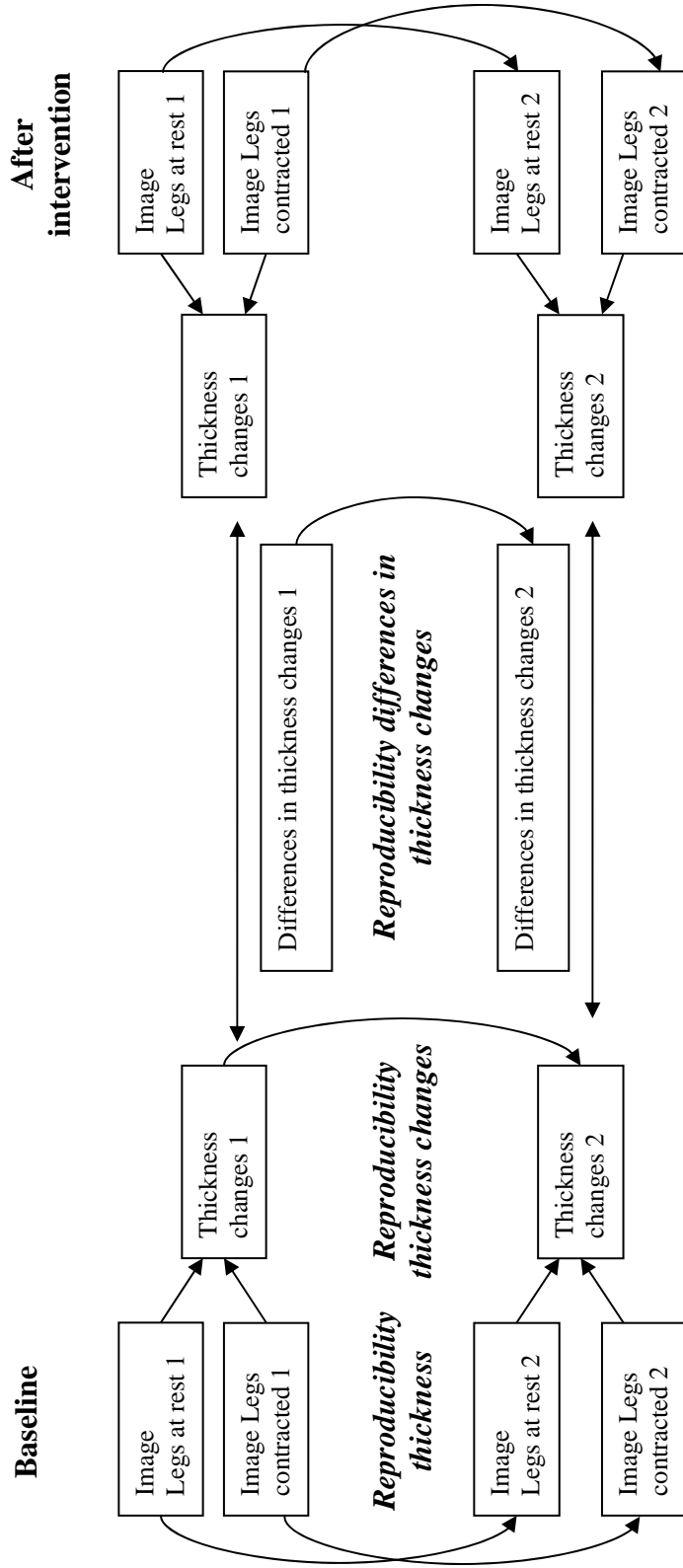


Figure 2. Diagram representing how the reproducibility tests were performed.

Chapter Six

The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments

Chapter Six is published as:

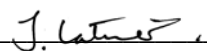
Maher CG, Latimer J, Costa LOP (2007) The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments. *Revista Brasileira de Fisioterapia* 11:245-252.

**Statement from co-authors confirming authorship contribution of the
PhD candidate**

As co-authors of the paper “The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments”, we confirm that Leonardo Costa has made the following contributions:

- Conception and design of the research
- Writing of the manuscript and critical appraisal of the content

Christopher G Maher _____  _____ Date 20.03.2009

Jane Latimer _____  _____ Date 20.03.2009

THE RELEVANCE OF CROSS-CULTURAL ADAPTATION AND CLINIMETRICS FOR PHYSICAL THERAPY INSTRUMENTS

MAHER CG, LATIMER J & COSTA LOP

Back Pain Research Group, The University of Sydney, Sydney, NSW - Australia

Correspondence to: Leonardo O. P. Costa, Back Pain Research Group, The University of Sydney, East Street Lidcombe, Sydney NSW Australia 2141, e-mail: lcos3060@usyd.edu.au

Received: 26/07/2007 - Revised: 28/07/2007 - Accepted: 01/08/2007

ABSTRACT

Background: Self-report outcome measures (questionnaires) are widely used by physiotherapists for measuring patient's health status or treatment outcomes. Most of these measurement tools were developed in English and their usefulness is very limited in non-English speaking countries such as Brazil. The only way to solve this problem is to properly adapt the relevant questionnaires into a target language and culture (e.g. Brazilian-Portuguese) and then test the instrument by checking its psychometric (clinimetric) characteristics. Objectives: The purpose of this paper was to present relevant issues in the process of cross-cultural adaptations and clinimetric testing for self-report outcome measurements. Advice on how to perform a cross-cultural adaptation, how to properly check the clinimetric properties, how to select a relevant questionnaire and how to evaluate the quality of an adapted questionnaire are provided. Additionally we present all Brazilian-Portuguese cross-cultural adaptations of low back pain measurements that we know of. Conclusions: There is a clear need for more effort in the field of cross-cultural adaptation and clinimetrics, without proper instruments, the management of patients from non-English speaking countries is compromised.

Key words: questionnaire; reliability; validity; psychometric properties; clinimetrics.

RESUMO

A importância da adaptação transcultural e clinimétrica para instrumentos de fisioterapia

Introdução: Questionários vem sendo amplamente utilizados por fisioterapeutas para medir a condição de saúde do paciente ou dos resultados de tratamento. A maioria desses instrumentos para avaliação foi desenvolvida em inglês, sendo seu uso bastante limitado em países que não usam o inglês como língua nativa, a exemplo do Brasil. A única forma de resolver esse problema é através de uma adaptação apropriada dos questionários relevantes para um alvo lingüístico e cultural (por exemplo, português do Brasil) e então testar suas características psicométricas (clinimétricas). Objetivo: A finalidade deste artigo foi a apresentar os tópicos relevantes no processo das adaptações transculturais de questionários e os seus respectivos testes clinimétricos. São fornecidas propostas sobre como realizar uma adaptação transcultural, como avaliar adequadamente as propriedades clinimétricas, como selecionar um questionário relevante e como avaliar a qualidade de um questionário adaptado. Além disso, são também apresentadas as adaptações conhecidas, para português do Brasil, dos questionários para avaliação de dor lombar. Conclusão: Existe uma clara necessidade de mais esforços na área de adaptação transcultural e clinimetria. Sem os instrumentos adequados, o gerenciamento no cuidado de pacientes nos países onde o inglês não é a língua nativa torna-se comprometido.

Palavras-chave: questionário; confiabilidade; validade; propriedades psicométricas; clinimetria.

INTRODUCTION

What do we mean by cross-cultural adaptation?

Physiotherapists commonly use questionnaires to assess treatment outcomes. These pen and paper tests provide a convenient way to assess how physical therapy treatment has affected outcomes such as pain, disability and health-related quality of life. Questionnaires can also be used by physiotherapists to screen for serious diseases such as cancer¹, to screen for co-morbidities that may be precautions to exercise², to estimate prognosis³ and to collect information on how their patient is thinking or feeling⁴. For a questionnaire to be useful a patients needs to be able to read and understand the text and the items need to make sense and be relevant to that person.

It would be possible to use the one questionnaire across the globe if all the people of the world spoke the same language, they felt and thought the same way and their lifestyles were much the same. Thankfully the world is a much more interesting place than this and to accommodate this variety it is necessary for researchers to adapt the original questionnaire so that it is comprehensible and relevant in the new setting. This process is called cross-cultural adaptation.

Cross-cultural adaptation is more complex than it may first seem. It is a common, though questionable, practice to use an English language questionnaire for anyone who speaks English. One problem with this approach is that the best English word to express a concept may vary between different countries. For example there are separate Australian/New Zealand, Canadian, USA and United Kingdom English language versions of the SF36 quality of life questionnaire to accommodate the different ways that the English language is used in these countries⁵. In the Australian/New Zealand and UK versions, item 9a, used to assess vitality, is "Did you feel full of life" whereas the Canadian and US versions use the wording "Did you feel full of pep". There are also differences in the use of imperial and metric units of measurement in the various SF36 versions to reflect the local customs. There are likely to be similar variations in the use of Portuguese language for the various countries where Portuguese is spoken. So where possible, Brazilian physiotherapists should use Brazilian-Portuguese questionnaires, not Portuguese versions.

Even within a country it would be simplistic to presume that a single questionnaire could be used for all people. In some countries like Canada and Belgium, there is more than one official language. Even in countries where there is only one official language, barriers to understanding may be present. For example in Australia although the official language is English there are over 200 languages spoken with ~16% of the population speaking a language other than English at home⁶. In the USA ~18% of the population⁷ speak a language other than English at home and of these about half have difficulty with English.

Even if the people of the world all spoke the same language a single assessment instrument would not be possible. This is because the people of the world do not share a common culture and lifestyle. As an illustration the Neck Disability Index was developed in North America to assess neck pain-related disability with two of the ten items assessing how neck pain limits recreation and driving⁸. The potential problem is that not all the world partakes in these activities. Similarly the Shoulder Disability Questionnaire⁹ was developed in Europe and includes the item "Because of my shoulder I have trouble putting on a coat or sweater" which is probably not relevant in tropical countries.

Why do we need cross-cultural adaptations of outcome measures?

Most questionnaires that a physiotherapist would use were originally developed in the English language. Cross cultural adaptation of existing English language questionnaires would enable comparisons of different populations and permit the exchange of information across cultural and linguistic barriers. The ready availability of adapted questionnaires would also stop the common, though undesirable practice, of excluding subjects from clinical trials on the grounds that they do not understand the language spoken in that country.

An important reason to adapt an existing questionnaire, is that it is much more efficient than developing a new one. There is substantial work involved in developing and validating a questionnaire. As an illustration the items from the Quebec back pain disability questionnaire¹⁰ were selected based upon a series of studies surveying clinical experts and patients that identified over 130 potential items. After this preliminary stage psychometric testing (also called clinimetric testing) was undertaken to identify the final 20 items that comprise the questionnaire.

How does cross-cultural adaptation differ from translation?

Translation is not the same as cross-cultural adaptation. Translation may simply involve a single person translating the questionnaire from one language to another. In contrast cross-cultural adaptation involves a team including translators, health professionals who would normally use the questionnaire and researchers with an understanding of clinimetrics. The process of cross-cultural adaptation includes initial translation, synthesis, back translation, expert committee review, pilot testing of the draft translation and psychometric evaluation. The process is designed to achieve equivalence between the original and translated versions with respect to language (semantic and idiomatic equivalence) and also equivalent experiences of daily life and meaning/ways of thinking¹¹.

It is important to understand that the goal of cross-cultural adaptation is a questionnaire equivalent to the original; which is different from aiming to produce an identical questionnaire. There can be situations where a questionnaire

item is irrelevant or not meaningful in a new setting and it makes sense to adapt or replace the item. If this is done the revised item should be chosen to reflect the construct/attribute that was being measured with the original item and it may be useful to consult the developer of the original test.

There are problems with using questionnaires that have only been translated because translation is only one of the steps involved in cross-cultural adaptation^{11,12}. Mistranslations are possible in translated questionnaires because the full process of cross-cultural adaptation includes additional steps designed to identify mistranslation. Perhaps surprisingly, overly literal translation may cause even more problems. Trying to make the translation exactly mimic the structure and ordering of words of the original text can provide a translation that is awkward to read because the syntax of the two languages may be quite different. Literal translation may also create text that makes no sense or has a very different meaning to that of the original. As an illustration there are idiomatic expressions in English language that have a meaning quite different to that normally conveyed by the component words. For example the wording from the DASS (Depression Anxiety Stress Scale) "I felt down-hearted and blue" (feeling depressed) and the item from the MHLOC (Multidimensional Health Locus of Control) "If my condition takes a turn for the worse..." (if my condition deteriorates). Another possibility is that the translated words may be understood but irrelevant in the new setting. Lastly there may also be English words that cannot be directly translated because there is no equivalent word in the target language. These potential problems are avoided with cross-cultural adaptation.

Guidelines for cross-cultural adaptation (advice for researchers)

The process of cross-cultural adaptation includes initial translation, synthesis, back translation, expert committee review, pilot testing of draft translation and psychometric evaluation¹¹⁻¹³. To illustrate the process we will presume that an English language questionnaire, originally developed in Australia, is being adapted for use in Brazil.

Guidelines for cross-cultural adaptation advise that two or more translators should independently translate the source questionnaire¹¹⁻¹³. Multiple translations permit the identification of errors or misinterpretations in translation that would not be apparent with one translation. It has been suggested that the translators should be chosen so that they have different backgrounds and include translators who do, and do not, have an understanding of the concepts being assessed in the questionnaire¹¹. It has also been advised that the translators should be translating into their mother tongue^{11,12} and preferably be both bilingual (speak both English and Portuguese) and bicultural (have lived in Australia and Brazil)¹².

Following translation from English to Portuguese the translators and an observer synthesise the multiple translations to produce a consensus Portuguese translation. Translators

blind to the original questionnaire then translate the consensus Portuguese translation back into English. As before, multiple translators who are bilingual and bicultural are preferred but for the back translation their mother tongue should be English. Beaton et al.¹¹ advise that the back translators should be naïve to the concepts involved in the original questionnaire whereas van Widenfelt et al.¹² advise that they should have some knowledge of the area. Given there are arguments for both suggestions a sensible position would be to include both types of back translators.

An expert committee then reviews the original English language questionnaire, the Brazilian translations, the consensus Brazilian translation, the English back translations and any notes taken. The committee should comprise the translators, health professionals who would normally use the questionnaire and researchers with an understanding of clinimetrics. It can also be useful to include the developers of the original questionnaire. The goal is to produce a preliminary Brazilian-Portuguese version of the questionnaire that is equivalent to the original questionnaire. This preliminary questionnaire then undergoes pilot testing with members of the target population. Subjects from a range of backgrounds are asked to complete the questionnaire and then are interviewed by the researchers. This gives an opportunity to establish that the people who will complete the questionnaire are interpreting it in the same way as the expert committee have. Lastly the new instrument needs to undergo psychometric testing because it cannot be presumed that the test properties of the English language version will apply to the adapted questionnaire.

Psychometric (Clinimetric) Properties

The aim of the process of cross cultural adaptation is to produce semantic and idiomatic equivalence between the source and the target languages/cultures. When we perform a cross-cultural adaptation of a questionnaire we assume that these procedures will retain the original psychometric properties of the questionnaire, however this assumption may not hold because there may be cultural differences between the two populations. Because it is crucial to test the psychometric properties of the adapted questionnaire in the target population after the cross-cultural adaptation procedures, a brief explanation of the psychometric concepts are described below^{14,15}.

Internal Consistency

A good questionnaire must be homogeneous; the items should be addressing different aspects of the same construct. For example the Roland Morris Disability Questionnaire¹⁶ has 24 items that address disability in low back pain patients, every single item from this questionnaire refers to a different activity but all items are related to disability in general. Most questionnaires measure a single underlying construct by using multiple items, and these items should be moderately correlated with each other, and each item should correlate with the total

scale score; these two factors form the basis of the *Internal Consistency* of the scale¹⁴. Internal consistency can be evaluated by calculating Cronbach's alpha. A low alpha value means that some items are measuring other constructs, while a very high Cronbach's alpha means that the items in the questionnaire show too much homogeneity and some items may be redundant. Current guidelines also suggest performing a factor analysis in order to confirm the internal consistency¹⁵.

Reproducibility

Reproducibility is the extent to which repeated measurement on stable subjects yields similar results¹⁷. Reproducibility comprises two related constructs: agreement and reliability. Agreement statistics describe how close the scores for repeated measures are whereas reliability statistics describe the correlation between repeated measures. With reproducible instruments clinicians and researchers will be able to draw conclusions satisfactorily, formulate theories or make claims about generalizability. It is useful to check if the measurement tool is reproducible on different occasions (intra-tester reproducibility) and with different assessors (inter-tester reproducibility). It is usually easy to interpret reliability statistics as most are expressed on a scale from 0 to 1, where zero indicates no reliability and 1 indicates perfect reliability.

Validity

In the previous paragraph we examine how reproducible a test is under different conditions. The next step is to check if the measurement tool is assessing the specific construct that it was developed, that is *validity*. We can test validity by correlating the scores of a measurement with a *gold standard* (criterion validity), however various constructs that physiotherapists assesses in clinical practice such as health-related quality of life, pain or disability have no gold standard, in this case we can test validity by correlating the scores with another tool that measures the same construct (construct validity).

Responsiveness

We can easily assume that the goal for any kind of treatment is to induce changes in patient's health status and the *responsiveness* is the ability of a questionnaire to detect clinically important changes over time, even if these changes are small¹⁵. There are two approaches to measuring responsiveness. The more typical approach is to study patients where true improvement is expected and then calculate the effect size (which is the ratio of the mean difference to the standard deviation at baseline). Another approach is to use an external criterion of true change and investigate how well the measure can discriminate between subjects who have truly improved and those who did not¹⁵.

Ceiling and floor effects

Imagine a 10-item disability questionnaire, where each item is scored from zero (meaning no disability with that item) and ten (meaning completely unable to perform that item) and the total score is the sum of the item scores. The questionnaire is unable to detect deterioration in patients who score the maximum score (100) or improvement in patients who scored the lowest score (0). *Floor or ceiling effects* are considered to be present if more than 15% of respondents achieved the lowest (floor effect) or highest possible score (ceiling effect), respectively¹⁵. Moreover ceiling and floor effects have clear implications on the reproducibility and responsiveness of the questionnaire.

How to properly test a questionnaire

Testing a questionnaire is a very time-consuming task; we first need to identify a relevant and generalizable sample for testing. A recently published guideline for evaluating measurement properties of health status questionnaires recommends that at least 100 patients are necessary to analyse all psychometric properties¹⁵. Preferably the patients should be under treatment and all patients should answer the questionnaire(s) on three occasions: the first one at baseline (in this step it is possible to calculate validity, internal consistency and ceiling and floor effects), the second testing occasion should be chosen so that the construct being measured would not be expected to have changed (for example, we do not expect an important change in pain and disability scores in chronic low back pain (LBP) patients in 24 hours); this data will be useful for analysing test-retest reproducibility. Finally the third testing occasion should be selected so that it is likely that true change has occurred in the patient's status (for example, we expect that patients with acute LBP should improve in 2-3 weeks with physical therapy treatment). This third testing occasion allows for the assessment of responsiveness of the instrument (see figure below). It is useful to include a global change scale (such as the Global Perceived Effect Scale), by using it we can establish whether the patients have changed or not over the follow up periods.

HOW TO SELECT A CROSS-CULTURAL ADAPTATION: (ADVICE FOR CLINICIANS AND RESEARCHERS WISHING TO USE A CROSS-CULTURAL ADAPTATION)

Searching for adapted questionnaires

You have decided on the patient outcomes you wish to measure and now need to find a relevant questionnaire in Brazilian-Portuguese. How do you go about doing this?

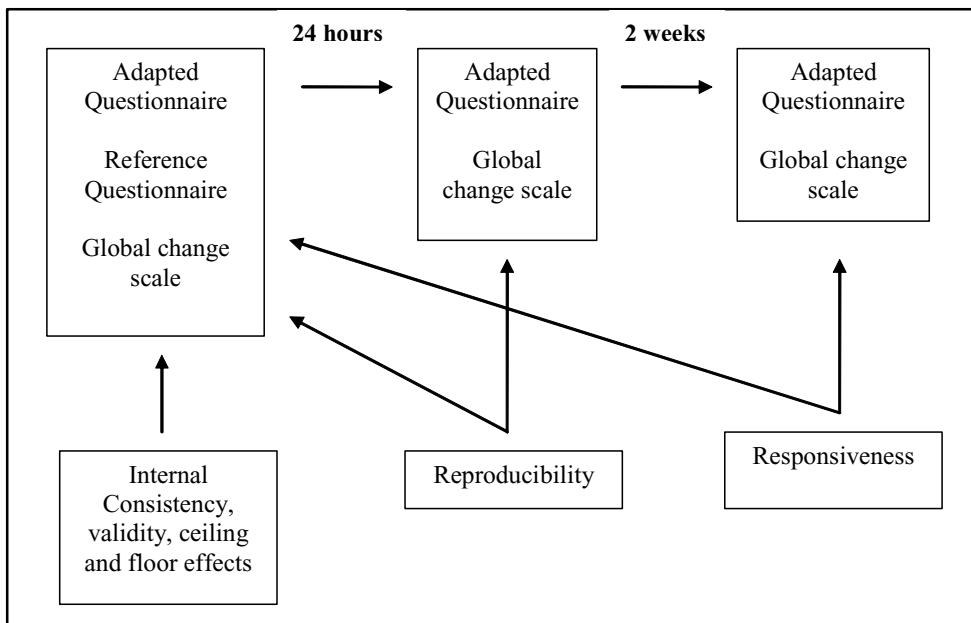


Figure 1. Example design for psychometric evaluation of an adapted questionnaire.

One strategy would be to search the internet for all relevant questionnaires that purport to measure your outcome of interest and then search for translated versions of these. If you are a researcher and subscribe to various medical databases you could follow a procedure similar to that used by Costa et al.¹⁸, in their search for international cross-cultural adaptations of self-report outcome measures for low back pain. For example, if you were interested in measuring disability in patients with low back pain you would perform two searches of the MEDLINE, EMBASE, CINAHL and LILACS databases.

The purpose of the first search would be to identify the low back pain questionnaire and could be performed by using search terms relevant to *low back pain AND questionnaire OR outcome measure*. To identify a cross-cultural adaptation, a second search would be performed where the name of the questionnaire identified in the first search would be used as the first search term (eg: *Roland Morris*) and then the terms *validation, translation, cross-cultural adaptation* plus the name of the desired language (eg: *Portuguese*) added. [The results of the search performed by Costa et al identified 40 relevant LBP self-report questionnaires of which fifteen had a cross-cultural adaptation. Only 19 of the 35 main languages that they searched for were represented in the search results, suggesting that while there are a large number of original questionnaires available, very few have been translated into other languages. Had a cross-cultural adaptation existed for each of the main languages they would have found 1400 possible adaptations (40 questionnaires x 35 languages). Clearly there is an urgent need for further cross-cultural adaptations to be conducted]¹⁸.

If you are a clinician without access to paid databases you could still search MEDLINE using PubMed which is freely available online. The PubMed URL is www.pubmed.gov.

Evaluating the quality of the located questionnaires

By now you have located a translated version of your questionnaire on the internet and are deciding whether to use it. You need to consider that there are many translated and adapted measures available on the web, some of them of very poor quality, and it is therefore important to exercise caution when deciding which measure to select. The best translated measures will be derived from peer-reviewed sources and will have followed guidelines for cross-cultural adaptation as described in the previous section. Also you need to consider whether appropriate psychometric testing of the adapted questionnaire was performed.

There are several scales available that rate the quality of psychometric testing of health questionnaires^{14,15}. The important quality criteria include satisfactory evaluation of:

- Internal consistency
- Construct validity
- Reproducibility
- Responsiveness
- Ceiling and floor effects

In order to determine the best adapted measure to use, a table is provided below with an item checklist and the criteria necessary to obtain a positive, indeterminate or negative rating (adapted from Terwee et al.¹⁵). The measure receiving the greatest number of positive ratings would be the best measure to use.

Table 1. Quality criteria for health status questionnaires.

Quality Criteria	Rating		
	+ = <i>positive</i>	? = <i>indeterminate</i>	- = <i>negative</i>
<i>Internal Consistency</i>	Factor analysis performed on adequate sample (≥ 100), Cronbach's alpha(s) between 0.70-0.95 for each dimension	No factor analysis OR doubtful design or method	Cronbach's alpha(s) < 0.70 or > 0.95 , despite adequate design and method
<i>Construct Validity</i>	Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses	doubtful design or method (e.g. no hypotheses)	less than 75% of hypotheses were confirmed, despite adequate design and methods
<i>Reproducibility</i>	ICC or Kappa ≥ 0.70	doubtful design or method (e.g. time interval not mentioned)	ICC or Kappa < 0.70 , despite adequate design and method
<i>Responsiveness</i>	Smallest detectable change _{individual} or Smallest detectable change _{group} $<$ Minimal important change OR Minimal important change outside the limits of agreement OR Responsiveness ratio > 1.96 OR Area under the curve ≥ 0.70	doubtful design or method OR sample size < 50 OR methodological flaws	Smallest detectable change _{individual} or Smallest detectable change _{group} \geq Minimal important change OR Minimal important change equals or inside limits of agreement OR Responsiveness ratio ≤ 1.96 OR Area under the curve < 0.70 , despite adequate design and methods
<i>Ceiling and floor effects</i>	$\leq 15\%$ of the respondents achieved the highest or lowest possible scores	doubtful design or method OR sample size < 50 OR methodological flaws	$> 15\%$ of the respondents achieved the highest or lowest possible scores, despite adequate design and methods

0 = no information available

no information found on internal consistency

no information found on construct validity

no information found on reproducibility

no information found on responsiveness

no information found on floor and ceiling effects

The study by Costa et al.¹⁸ revealed that the quality of the psychometric evaluation of the adapted low back pain questionnaires was typically poor with most of the testing confined to the evaluation of reproducibility and construct validity of the measure. There is clearly a need for more research in this area.

Will this adapted questionnaire be useful for my patients?

In some circumstances questionnaires will have undergone cross-cultural adaptation and high quality psychometric testing but may not be useful for your patients. For instance, the items of The Roland Morris Disability Questionnaire (RMDQ) provide an insight into how a patient's low back pain is impacting on their experience of daily life. However, some of the experiences listed in this questionnaire may not be the experiences of daily life in another country or culture. Consider a Thai version of the RMDQ¹⁹ that contains the Thai translation of the English item listed below:

16. I have trouble putting on my socks (or stockings) because of the pain in my back

This question may not be relevant to many patients with low back pain living in rural areas of Thailand where putting on socks (or stockings) is not part of the daily experience. This item may also be irrelevant to people living in parts of Africa or India.

A further example of irrelevant items occurs with The Functional Rating Index (FRI) which contains items that ask about driving. There are many places in the world where such an item would not be relevant to the patient being evaluated.

In summary, careful consideration needs to be given to the items contained in the located questionnaire to determine whether they will be relevant to that patient's experience of life.

AVAILABLE BRAZILIAN-PORTUGUESE CROSS-CULTURAL ADAPTATIONS, AN EXAMPLE FROM LBP QUESTIONNAIRES

A recent systematic review performed by our research group on cross cultural adaptations for self-report outcome measures for LBP patients stated that "there is a clear need for further cross-cultural adaptation of LBP self report measures and a great attention to qualify of psychometric evaluation of adapted questionnaires"¹⁸. By the time of publication, there was just one questionnaire adapted in Brazilian-Portuguese. We have updated the data from search strategies recently and so far we have found just three questionnaires relevant to the management of LBP that are available for Brazilians.

The first adaptation in Brazilian-Portuguese was performed with the Roland Morris Disability Questionnaire (RMDQ)¹⁶, the RMDQ is a self-report questionnaire consisting of 24 items related to normal activities of daily living. The RMDQ was developed by selecting 24 relevant items from

the longer Sickness Impact Profile²⁰. The questionnaire was transformed to a condition-specific measure for back pain by adding the phrase "because of my back" to each statement. The patients are asked to circle those items, which they perceive as difficult to perform due to back pain. Each answer is scaled simply 0 or 1, thus leaving a range of scores of 0 to 24, a higher score indicating higher disability. There are two studies that tested the RMDQ in Brazilian-Portuguese, the first one²¹ tested the questionnaire in 30 Brazilian patients with LBP for reliability (ICC 0.94) and construct validity with the pain visual analogue scale (Pearson's r: 0.76 p< 0.01) and with pain numerical scale (0-6) (Pearson's r: 0.80 p< 0.01).

This questionnaire was retested by Costa et al.²² in 140 Brazilian LBP patients finding a very high level of reliability [ICC 0.95 (95%CI 0.93-0.97)] and internal consistency (Cronbach's alpha 0.92). A high correlation between the Brazil Roland Morris and the Functional Rating Index²³ was observed (Pearson's r: 0.80 p<0.001) and the RMDQ correlated moderately with the Pain Numerical Rating Scale at baseline (Pearson's r: 0.55 p<0.001) showing good construct validity. The authors did not find any ceiling or floor effects however the responsiveness was quite small [Effect size 0.10 (84% CI 0.04-0.16)].

The second adaptation in Brazilian-Portuguese was the Oswestry Disability Index (ODI)²⁴. The ODI measures self perceived disability in LBP patients, it is a 10-item questionnaire with each item scored on a 0-5 *Likert* scale. The ODI is scored by summing the item responses and expressing the total as percentage of the maximum score. If the patient fails to complete a section, the percentage score is adjusted. The total ODI score ranges from 0 (no disability) to 100 (maximal disability). The Brazilian-Portuguese version of the ODI²⁵ was tested in 120 LBP patients, the authors finding good internal consistency (Cronbach's alpha 0.87) and excellent reliability (ICC 0.99). There was evidence of construct validity as the ODI was moderately correlated with pain measurement (Pearson's r: 0.66) and highly correlated with the Brazilian version of the RMDQ (Pearson's r: 0.81).

The third LBP questionnaire tested in Brazil was the Functional Rating Index (FRI)²³ this questionnaire was developed to measure perceived disability in patients with back and/or neck pain²³. The FRI emphasises function while concurrently measuring the patient's opinion, attitude, and self-rating of disability. The FRI is shorter than the RMDQ and can be used for back and neck pain. The Brazilian-Portuguese version of the FRI²² was tested in 140 patients and a very high level of reliability [ICC 0.95 (95%CI 0.93-0.97)] and internal consistency (Cronbach's alpha 0.92) was found. A high correlation with the RMDQ was observed (Pearson's r: 0.80 p< 0.001) and the FRI correlated moderately with the Pain Numerical Rating Scale at baseline (Pearson's r: 0.67 p< 0.001) showing good construct validity. The authors did not find any ceiling or floor effects and the

responsiveness was small [Effect size 0.18 (84% CI 0.11-0.24)].

Directions for the future

Our recent experience of searching for international cross-cultural adaptations identified:

(i) the urgent need for more cross-cultural adaptations of useful self report outcome measures

(ii) the need for high quality psychometric evaluation of such measures

(iii) the importance of publishing the adaptation, with an additional English abstract if possible, in a journal that is indexed and therefore included in a freely available database such as *PubMed*

(iv) the need for a repository where all cross-cultural adaptations and their psychometric evaluations could be stored and then accessed by interested researchers and clinicians freely on-line. Beaton et al.¹¹ advocated this in their original publication outlining the process for cross-cultural adaptation, but to date such a repository has not been developed. It would appear most efficient if such a repository was developed and maintained by an internationally recognised, non-discipline specific professional organisation such as the Cochrane Collaboration. Such a repository would help ensure that multiple translations are not in use at the one time, and would prevent the costly and time-consuming task of replicating an already well adapted version of a questionnaire.

Acknowledgements: Chris Maher holds a fellowship funded by the National Health Medical Research Council, Australia. Leonardo Costa is a PhD supported by CAPES, Ministério da Educação, Brazil and by the Pontificia Universidade Católica de Minas Gerais, Brazil.

REFERENCES

- Delitto A, Erhard R, Bowling R. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther.* 1995;75: 470-89.
- Whaley M, Brubaker P, Otto R. ACSM's guidelines for exercise testing and prescription. Philadelphia: Lippincott Williams & Wilkins; 2006.
- Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain.* 1998;14:209-15.
- Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A fear-avoidance beliefs questionnaire (FABQ) and the role of fear avoidance beliefs in chronic low back pain and disability. *Pain.* 1993;52(2):157-68.
- Trust MO. SF-36 Health survey scoring manual for english language applications: Australia/New Zealand, Canada, United Kingdom. Boston: Medical Outcomes Trust; 1994.
- Australian Bureau of Statistics. Canberra: Australian Bureau of Statistics, 2006.
- US Bureau of Statistics. Washington: US bureau of Statistics; 2000.
- Vernon H, Mior S. The neck disability index: A study of reliability and validity. *J Manip Physiol Ther.* 1991;14:409-15.
- van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. *J Clin Epidemiol.* 2000;53(1):29-38.
- Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale: Conceptualisation and development. *J Clin Epidemiol.* 1996;49(2):151-61.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186-91.
- van Widenfelt BM, Treffers PD, de Beurs E, Siebelink BM, Koudijs E. Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clin Child Fam Psychol Rev.* 2005;8(2):135-47.
- Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol.* 1993;46:1417-32.
- Streiner DL, Norman GR. Health Measurement Scales: a practical guide to their development and use. New York: Oxford Medical Publications; 2003.
- Terwee CB, Bot SD, de Boer MR, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34-42.
- Roland M, Morris R. A study of natural history of back pain, part 1: development of a reliable and sensitive measure of disability in low back pain. *Spine.* 1983;8:141-4.
- de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59:1033-9.
- Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. *Spine.* 2007;32:1028-37.
- Jirattanaphochai K, Jung S, Sumanant C, Saengnipanthkul S. Reliability of the Roland - Morris Disability Questionnaire (Thai version) for the evaluation of low back pain patients. *J Med Assoc Thai.* 2005;88(3):407-11.
- Bergner M, Bobbitt RA, Carter WB. The sickness impact profile: development and final revision of a health status measure. *Med Care.* 1981;19:787-805.
- Nusbaum L, Natour J, Ferraz MB, Goldenberg J. Translation, adaptation and validation of the Roland-Morris questionnaire-Brazil Roland-Morris. *Braz J Med Biol Res.* 2001;34(2): 203-10.
- Costa LOP, Maher CG, Latimer J, Ferreira PH, Pozzi G, Ribeiro R. Psychometric Characteristics of the Brazilian-Portuguese Versions of the Functional Rating Index and the Roland Morris Disability Questionnaire. *Spine* 2007, 32(17):1902-1907.
- Feise RJ, Menke MJ. Functional rating index - a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine.* 2001;26:78-87.
- Fairbank JCT, Couper J, Davies JB. The Oswestry low back pain disability questionnaire. *Physiotherapy.* 1980;66:271-3.
- Vigatto R, Alexandre NM, Correa Filho HR. Development of a brazilian portuguese version of the oswestry disability index: cross-cultural adaptation, reliability, and validity. *Spine.* 2007;32:481-6.

Chapter Seven

Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations

Chapter Seven is published as:

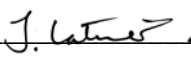
Costa LOP, Maher CG, Latimer J (2007) Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations. *Spine* 32:1028-1037.

**Statement from co-authors confirming authorship contribution of the
PhD candidate**

As co-authors of the paper “Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations”, we confirm that Leonardo Costa has made the following contributions:

- Conception and design of the research
- Data collection
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

Christopher G Maher _____  _____ Date 20.03.2009

Jane Latimer _____  _____ Date 20.03.2009

Self-Report Outcome Measures for Low Back Pain

Searching for International Cross-Cultural Adaptations

Leonardo Oliveira Pena Costa, PT, MSc, Chris G. Maher, PT, PhD, and Jane Latimer, PT, PhD

Study Design. Systematic review.

Objectives. To describe the available cross-cultural adaptations of low back pain (LBP) self-report outcome measures and the psychometric testing that has occurred for each adaptation.

Summary of Background Data. Self-report measures are commonly used in clinical practice and in research studies. Most existing questionnaires were developed in English, and it is not clear how many have been adapted to other languages.

Methods. Two different searches on MEDLINE, EMBASE, CINAHL, and LILACS were performed. The first search identified questionnaires specifically designed for patients with LBP. The second search combined the name of the questionnaire with 35 different languages in order to locate cross-cultural adaptations of the questionnaire. Data on the psychometric testing of the translated questionnaires were extracted.

Results. Forty questionnaires were identified, only 15 of which had been adapted to a new language. Only 19 of the 35 different languages we searched for were represented in the search results. From 1400 possible adaptations, only 61 have been completed. Psychometric testing of the adapted questionnaires was quite variable and in general suboptimal with testing usually restricted to an assessment of reliability and construct validity.

Conclusions. There is a clear need for further cross-cultural adaptation of LBP self-report measures and for greater attention to the quality of psychometric evaluation of adapted questionnaires. Without appropriately adapted measures, the clinical management of LBP patients who do not speak English is potentially compromised.

Key words: outcome assessment, psychometric properties, low back pain. *Spine* 2007;32:1028–1037

Self-report outcome measures are commonly used in clinical practice, in clinical research and large epidemiologic studies. Many are simple, reliable, and of low cost, making them suitable for quality assurance and research

activities. Most existing self-report measures were developed in English, and it is not clear how many have been adapted to other languages.

Cross-cultural adaptation of existing self-report measures is important for a number of reasons. First, not all the world speaks English; and even in English-speaking countries like Australia, United Kingdom, and United States, a significant proportion of the population are not English-speakers. Second, the availability of adapted questionnaires may stop the undesirable, but common, practice of excluding non-English subjects from clinical trials conducted in English-speaking countries. Third, the existence of adapted questionnaires would be of value to researchers conducting systematic reviews by assisting the pooling of data from trials conducted in non-English-speaking countries. Lastly, adaptation of existing questionnaires is potentially more simple and efficient than requiring researchers in non-English-speaking countries to develop their own self-report measures.

Cross-cultural adaptation is more than just simple translation of English text. Researchers need to follow guidelines for cross-cultural adaptation to ensure that the adapted questionnaire is semantically equivalent to the original and that the items and scale are relevant in the new culture. Following this first step, it is also essential to assess whether the adapted questionnaire has retained the content validity of the original questionnaire. This second step requires an assessment of the measurement properties of the new questionnaire.¹

The measurement properties (or psychometric properties) of a questionnaire include both reliability and validity of the calculated scores. A brief explanation of the concepts in psychometric testing is provided in Table 1.

The objectives of this paper are to describe the cross-cultural adaptations of self-report measures relevant to the management of low back pain (LBP) that are now available and describe the psychometric testing that has occurred for each adaptation.

■ Materials and Methods

Literature Search 1. The purpose of this first step was to identify self-report outcome measures designed for patients with LBP. A systematic literature search was carried out in MEDLINE, EMBASE, LILACS, and CINAHL databases for the period from January 2001 to July 2006. The following search terms were included: *back pain, low back pain, scoliosis, spinal stenosis, disc herniation, nerve root compromise, and ankylosing spondylitis* these terms were combined with the terms *questionnaire(s), outcome measure(s), index, and scale(s)*. The search strategy was: (back pain OR low back pain OR scoliosis OR spinal stenosis, OR disc herniation OR ankylosing spondylitis) AND (questionnaire(s) OR outcome mea-

From the Back Pain Research Group, School of Physiotherapy, University of Sydney, Lidcombe, NSW Australia.

Acknowledgment date: April 13, 2006. First revision date: July 17, 2006. Second revision date: August 11, 2006. Acceptance date: August 15, 2006.

Leonardo Costa is a PhD student supported by CAPES, Ministério da Educação, Brazil; Christopher Maher holds a research fellowship funded by the National Health and Medical Research Council of Australia.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Leonardo Oliveira Pena Costa, PT, MSc, Back Pain Research Group, School of Physiotherapy, University of Sydney, East Street, Lidcombe, NSW Australia 2141; E-mail: lcos3060@usyd.edu.au

Table 1. Concepts in Psychometric Testing¹⁸

Concept	Explanation
Internal consistency	Internal consistency is the extent to which the items assess the same construct. If a questionnaire has a no. of items addressing the same underlying dimension (e.g., disability status in low back pain patients), then it is reasonable to expect that scores on each item would be correlated with scores on all other items. Internal consistency can be determined by calculating Cronbach's alpha. A low alpha value means that some items may be ambiguous, while a very high Cronbach's alpha means that the items in the questionnaire show too much homogeneity and some items may be redundant
Factor analysis	Factor analysis confirms the structure of the questionnaire by summarizing patterns of correlations among observed variables, to reduce a large no. of items into a smaller no. of factors. Items that do not load on any factor or that load on multiple factors should be omitted or rephrased
Reliability	Reliability is the extent to which an experiment, test, or any measuring procedure yields the same result on repeated trials. Without the agreement of independent observers able to replicate research procedures or the ability to use research tools and procedures that yield consistent measurements, researchers would be unable to satisfactorily draw conclusions, formulate theories, or make claims about the generalizability of their research. A common measure of reliability is the intraclass correlation coefficient (ICC). This measure takes random error as well as systematic error (bias) into account. Other reliability indexes such as Pearson's correlation and coefficient of variation can be used
Responsiveness	Responsiveness is the ability of an instrument to measure real or important change over time, in the concept being measured
Validity	Validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. <i>Construct</i> validity refers to the extent to which scores on a particular instrument relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured. Where an external gold standard is available, to which the self-report measure could be compared, evaluation of <i>criterion</i> validity is possible. However, many self-report measures assess attributes such as quality of life or disability where there is no accepted external gold standard. In these cases, only assessment of <i>construct</i> validity is possible
Rasch analysis	Rasch analysis addresses unidimensionality by assessing the contribution that items make to the scale hierarchy. The technique also provides an estimate of item difficulty based on the frequency with which patients respond to an item, which can be used to assess the position of items along the scale and to consider any possible redundancy or gaps in the scale hierarchy
Floor or ceiling effects	Floor or ceiling effects are considered to be present if more than 15% of respondents achieved the lowest or highest possible score, respectively. If floor or ceiling effects are present, it is likely that items assessing the extremes of the attribute are missing from the scale. The consequence of a floor effect is that deterioration may be missed and for a ceiling effect improvement may be missed

sure(s) OR index OR scale(s)). In addition, hand searches of journals and textbooks of spinal disorders were carried out. Titles and abstracts of papers from the search were scanned to determine eligibility. The inclusion criteria were: 1) that the questionnaire had the word “back” in the text or in the title and 2) the questionnaire was specifically developed for LBP patients. Papers published in non-English journals were only included if they had an English abstract.

Literature Search 2. The purpose of this step was to locate different language versions available for the questionnaires identified in Step 1. The list from the first search was used for the second literature search, the terms used were the *questionnaire's name* (e.g., Roland Morris) and *validation, translation, cross-cultural adaptation* plus 35 different languages (e.g., Portuguese) using AND/OR operands to combine the terms (e.g., Roland Morris AND (validation OR translation OR cross-cultural adaptation) AND Portuguese). This second literature search was performed on the same databases used in the first search. There was no time limit on this search strategy.

Following the searches the list of questionnaires and cross-cultural adaptations was presented to 27 experts in spinal disorders from the United States, France, Canada, Netherlands, Germany, Norway, China, Tunisia, Turkey, Thailand, Italy, Denmark, Hong Kong, Spain, Switzerland, Korea, Japan, Greece, Brazil, Sweden, and Finland to determine whether the list was comprehensive.

Description of the Psychometric Testing of Translated Questionnaires. The following data were extracted from articles describing psychometric testing of the translated questionnaires: citation, sample size, and all types of psychometric properties possible (i.e., test-retest reliability, internal consistency, Rasch analysis, responsiveness, factor analysis, and construct validity). Additionally, psychometric properties were rated by the Quality Criteria for Psychometric Properties of Health Status Questionnaires² with the evaluation restricted to the subset of items relevant to cross-cultural adaptation (Appendix available online only through Article Plus). This checklist considers both the quality of psychometric testing and the results of psycho-

metric testing and so is somewhat different from scales used to measure the methodologic quality of trials. We are unaware of a scale that only measures quality of psychometric testing.

■ Results

From the first search, 66 potentially eligible questionnaires were identified from MEDLINE, 12 from CINAHL, 23 from EMBASE, and 1 from LILACS; however, neither CINAHL, nor EMBASE nor LILACS added new papers to the 66 obtained from MEDLINE. Twenty eight questionnaires did not meet inclusion criteria leaving 38 eligible questionnaires. The most common reasons for exclusion were the absence of the word “back” in the title or in the text, followed by no evidence that the questionnaire was specifically designed for LBP patients. The expert committee's review of the list of questionnaires found the list to be quite comprehensive. Committee members added 2 additional eligible questionnaires^{3,4} that were not found by the original searches.

From the second search in MEDLINE, 57 cross-cultural adaptation studies were identified. Only 15 of 40 eligible questionnaires had been adapted to a new language, and only 19 of 35 different languages we searched for were represented in the search results. As with the first search, the other databases did not add new international validations to the MEDLINE results. The experts found 4 new cross-cultural adaptations, providing a total of 61 cross-cultural adaptations. Of the 4 new adaptations, 1 was published in a nonindexed journal,⁵ 1 was an article “in press,”⁶ and 2 were from a paper that did not mention any of the terms that we used in our search.^{7,8}

Table 2 describes the 40 original LBP questionnaires and 61 cross-cultural adaptations. From 1400 possible translations (40 questionnaires by 35 languages), only 61 have been completed. The most frequently adapted

Table 2. LBP Questionnaires and Their International Adaptations

	Original Language	Arabic	Cantonese	Mandarin	Danish	Dutch	French	German
Acute Low Back Pain Screening Questionnaire ¹⁹	English							
Back Dysfunction Score ²⁹	English							
Back Beliefs Questionnaire ²²	English							
Back Illness Pain and Disability 9-Item Scale ²³	English							
Back Pain Functional Scale ²⁴	English							
Back Pain Interference Scales ²⁵	English							
Back Performance Scale ³	English							
Back-specific version of the SF-36 Physical Functioning Scale ²⁶	English							
Bath Ankylosing Spondylitis Disease Activity Index ²⁷	English							
Bath Ankylosing Spondylitis Functional Index ³¹	English							*14
Bournemouth Questionnaire ³⁵	English				*36			
CORE outcome measure ³⁷	English							
Core set ⁴	German							
Curtin Back Screening Questionnaire ³⁸	English							
Dallas Pain Questionnaire ³⁹	English						*40	
Dougados Functional Index for Ankylosing Spondylitis ⁴¹	English							*14
Fear Avoidance Beliefs Questionnaire ⁴⁴	English						*45	*46
Functional Outcomes Questionnaire: Spinal Disorders ⁴⁸	English							
Functional Rating Index ⁴⁹	English							
General Function Score ⁵⁰	English							
Hannover Functional Ability Questionnaire ⁵¹	German							
Low Back Outcome Score ⁵²	English							
Low Back Outcome Score for Back Pain ⁵³	English							
Low Back Pain Rating Scale ⁵⁴	English							*55
MODEMS Pain and Disability Lumbar Scale ⁵⁶	English							
NASS Lumbar Spine Outcome Assessment Instrument ⁵⁷	English							*13
Occupational Role Questionnaire ⁶⁰	English							
Outcome Measure for Lumbar Spinal Stenosis ^{61,62}	Swiss							
Oswestry Disability Index ⁵³	English	*12		*64	*6			*65
Perception of Disability Scale ⁷¹	English							
Quebec Back Pain Disability Scale ⁷²	English					*73	*74	
Roland Morris Disability Questionnaire ⁷⁶	English	*77	*78	*10	*9	*79	*11	*80
Scoliosis Research Society 22 ⁸⁷	English							
Spinal Function Sort ⁹²	English							
Spinal Pain Independence Measure ⁹³	Hebrew							
Aberdeen Low Back Pain Disability Scale ⁹⁴	English		*7					*95
The Maine-Seattle Back Questionnaire ⁹⁶	English							
Vermont Disability Prediction Questionnaire ⁹⁷	English							
Waddell Disability Index ⁹⁸	English							
Walter Reed Visual Assessment Scale ⁹⁹	English							

questionnaire was the Roland Morris Disability Questionnaire, which was adapted to 17 different languages, followed by the Oswestry Disability Index (11 languages), the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index for Ankylosing Spondylitis (5 languages), the Quebec Back Pain Disability Scale (4 languages), the Bath Ankylosing Spondylitis Disease Activity Index, the NASS Lumbar Spine Outcome Assessment Instrument, the Fear Avoidance Beliefs Questionnaire, and the Scoliosis Research Society (3 languages), the Aberdeen Low Back Pain Disability Scale (2 languages), and finally the Acute Low Back pain Screening Questionnaire, the Bournemouth, the Core-set, Low Back Pain Rating Scale, and the Dallas Pain Questionnaire with only one adaptation each

The most numerous languages in the cross-cultural adaptations were Spanish and German (8 times), followed by Turkish (7 times), French and Italian (4 times), Danish, Japanese, Norwegian, Persian and Swedish (3 times), Arabic, Mandarin, Cantonese, Dutch, Greek, and Thai (twice), and the other languages were only in one adaptation.

Table 3 summarizes the reported psychometric testing of the adapted questionnaires. There are large differences in the sample sizes used in psychometric testing ranging from 30 patients (Brazilian Portuguese and Dutch Ro-

land Morris) to 697 patients for the psychometric testing of the Greek Oswestry.

Almost all adaptations have been evaluated for test-retest reliability and internal consistency (48 of 61). Test-retest reliability is good to excellent with intraclass correlation coefficient values ranging from 0.53 (Turkish Bath Ankylosing Spondylitis Disease Activity Index) to 0.98 (Arabic Roland Morris). Other statistics were less commonly used to describe test retest reliability *e.g.*, Pearson's *r* (German Roland Morris: $r = 0.82$; $P = 0.0001$) and coefficient of variation (Norwegian Oswestry and Roland Morris: 12% and 15%, respectively). Internal consistency is also good to excellent with Cronbach's alpha ranging from 0.57 (Japanese Scoliosis Research Society) to 0.96 (Spanish Scoliosis Research Society). For internal consistency, Pearson's *r* (French Quebec: 0.44–0.76), Spearman's Rho coefficient (German Low Back Pain Rating Scale: 0.98), and Kuder-Richardson 20 coefficient (Chinese Hong Kong Roland Morris: 0.86) were also used.

Only 19 of 61 adapted questionnaires have been tested for responsiveness. Convergent validity was assessed in 57 of 61 questionnaires by comparing questionnaire scores to other measures measuring similar constructs. In many cases, the comparison measure seemed quite similar, *e.g.*, comparison of the Greek Roland Mor-

Table 2. Continued

Greek	Italian	Korean	Japanese	Norwegian	Persian	Portuguese	Romanian	Spanish	Swedish	Thai	Turkish
				*20							
	*15						*32	*28 *28	*29		*30 *33,34
								*8			
	*15							*28 *47	*42		*33,43
	*58							*59			
*66		*17	*67	*68	*69					*5	*70
					*69						*75
*66	*81		*67 *88	*68	*69	*82		*83 *89,90	*84	*85	*86 *91

ris to the Greek Oswestry where both measures are LBP specific measures of disability. However, some comparisons were more dissimilar, *e.g.*, the comparison of the Korean Oswestry to a Korean pain measure and a Korean quality of life measure. Rasch analysis was only performed on the Turkish Roland Morris version, and factor analysis was only performed on 8 adaptations.

The quality and outcome of the psychometric testing of each adaptation are described in Table 4. In general, there is acceptable evidence for reliability and construct validity but not internal consistency, responsiveness, or floor or ceiling effects. While internal consistency has been widely evaluated, the evaluation has typically not been acceptable. The problem with responsiveness and floor or ceiling effects is that these concepts have not usually been tested.

Discussion

Our study aimed to describe the cross-cultural adaptations of self-report measures relevant to the management of LBP. A total of 61 cross-cultural adaptations in 19 different languages were identified. Of the 40 original language questionnaires, only 15 have been adapted, which shows clearly that more effort in this field is required. There were also considerable differences in the

psychometric testing that has been undertaken with each adaptation.

When we visited the Oswestry website (<http://www.orthosurg.org.uk/odi/>) and Roland Morris e-mail (mroland@man.ac.uk), we found more versions than we did in our database searches. We contacted all authors by e-mail and established that all additional versions had only been published as theses or conference abstracts. We did not include these papers because these questionnaires have not been through a peer-review process and their quality thus remains unclear. In our view, the unpublished description and evaluation of such questionnaires are not useful.

In our study, we identified papers in non-English journals only if they contained an English abstract with an English description of the results. Ten questionnaires (8 papers^{5,9-15}) fulfilled this criterion. It is possible, however, that there were questionnaires available in other languages that were missed by our search strategy. Another problem is that some questionnaires have different names in different languages, for example: the French RMDQ is known in France as the EIFEL questionnaire; and due to this practice, some additional cross-cultural adaptations may have been missed. To avoid this problem, we advise that,

Table 3. Psychometric Characteristics of LBP Questionnaires

Language	Sample Size	Test-Retest Reliability (ICC)	Internal Consistency (Cronbach's alpha)	Responsiveness	Construct Validity (Pearson correlation)
Acute Low Back Pain Screening Questionnaire					
Norwegian ²⁰	123	0.90	0.95		RMDQ: 0.46, age: 0.38
Bath Ankylosing Spondylitis Disease Activity Index					
Spanish ²⁸	144	0.74		Effect size: 1.6	General well-being: 0.7, Pain VAS: 0.53, Morning stiffness: 0.64
Swedish ²⁹	113			Sensitivity to change: $P < 0.05$	BASFI: 0.64
Turkish ³⁰	71	0.53–0.85	0.80		BASFI: 0.62 ($P < 0.001$), Physician's assessment: 0.44 ($P < 0.001$)
Bath Ankylosing Spondylitis Functional Index					
German ¹⁴	72	0.92	0.81	SRM: 0.46	Schober test: -0.30 ($P < 0.05$), Fingertip to floor: 0.40 ($P < 0.001$), Pain VAS: 0.38 ($P < 0.01$)
Italian ¹⁵	95	0.91	0.90	ROC curve: 0.90	Pain VAS: 0.73, BASDAI: 0.72
Romanian ³²	41	0.82	0.93		Dougados Index: 0.37 ($P < 0.05$), Pain VAS: 0.44 ($P < 0.01$)
Spanish ²⁸	144	0.68		Effect size: 1.2	Schober test: -0.4 , Occiput to wall: 0.38
Swedish ⁴²	113			Sensitivity to change: $P < 0.001$	BASDAI: 0.68
Turkish ¹⁰⁰	81	0.76–0.95	0.93		Schober test: 0.44 ($P < 0.001$), Fingertip to floor: 0.56 ($P < 0.001$), Occiput to wall: 0.53 ($P < 0.01$)
Bournemouth Questionnaire					
Danish ³⁶	118	0.96	0.88–0.89		RMDQ: 0.46–0.73, SF-36: 0.64–0.71
Dallas Pain Questionnaire					
French ⁴⁰	59	0.75	0.89–0.91	Sensitivity to change: $P < 0.001$	Pain VAS: 0.78 ($P < 0.001$)
Core set					
Spanish ⁸	131		0.91	Effect size: 0.91–6.15	
Dougados Functional Index for Ankylosing Spondylitis					
German ¹⁴	72	0.89	0.85	SRM: 0.33	Schober test: -0.19 (NS), Fingertip to floor: 0.32 ($P < 0.01$), Pain VAS: 0.22 (NS)
Italian ¹⁵	95	0.86	0.87	ROC curve: 0.82	Pain VAS: 0.53, BASDAI: 0.65
Spanish ²⁸	144	0.87		Effect size: 1.05	BASFI: 0.83, Schober test: -0.36
Turkish ⁴³	70	0.61–0.88	0.91	Sensitivity to change: NS	Schober test: -0.29 ($P < 0.05$), Fingertip to floor: 0.45 ($P < 0.001$), Occiput to wall: 0.38 ($P < 0.01$)
Fear Avoidance Beliefs Questionnaire					
French ⁴⁵	217	0.72–0.88		Effect size: 0.98–1.37	Pain VAS: 0.36, Quebec: 0.36, HADS: 0.29
German ⁴⁶	302	0.87	0.89		Pain VAS: 0.26 ($P < 0.002$), FFbH-R: 0.36 ($P < 0.002$)
Spanish ⁴⁷	209	0.96	0.93		Pain VAS: 0.398 ($P = 0.00$), RMDQ: 0.522 ($P = 0.00$), Quality of life physical: -0.361 ($P = 0.00$), Quality of life mental D1: -0.361 ($P = 0.00$)
Low Back Pain Rating Scale					
German ⁵⁵	126		0.95		RMDQ: 0.91, SF-36: -0.34 to -0.72
NASS Lumbar Spine Outcome Assessment Instrument					
German ¹³	56	0.82–0.89			SF-36: 0.83, FFbH-R 0.28, Pain VAS: 0.68
Italian ⁵⁸	74	0.89–0.92	0.87–0.90		SF-36: -0.54 , Pain VAS: -0.43 to -0.58
Spanish ⁵⁹	70	0.63–0.91	0.78–0.90	ROC curve: 0.74 -0.81 , Effect size: -0.05 – 2.02	SF-36: -0.82 , Pain VAS: -0.63
Oswestry Disability Index					
Arabic ¹²	80	0.98	0.70–0.76		Quebec: 0.86, Pain VAS: 0.57
Mandarin ⁶⁴	79	0.86	0.81		
Danish ⁶	191	0.91	0.88		RMDQ: 0.78 ($P < 0.01$), Low back pain rating scale: 0.68 ($P < 0.01$), SF-36: 0.75 ($P < 0.01$)
German ⁶⁵	100	0.96	0.90		Pain VAS: 0.78 ($P < 0.001$), RMDQ: 0.80 ($P < 0.001$)
Greek ⁶⁶	697		0.83		RMDQ: 0.729 ($P < 0.0005$)
Japanese ⁶⁷	97	0.93	0.83		RMDQ: 0.785 ($P < 0.01$), JOA: 0.647 ($P < 0.01$)
Korean ¹⁷	206	0.91	0.84		Pain VAS: 0.425 ($P = 0.0001$), WHOQOL-BREF: -0.09 (NS) to -0.48 ($P = 0.001$)
Norwegian ⁶⁸	105	0.88	0.94		RMDQ: 0.73–0.60, Pain VAS: 0.39–0.52
Persian ⁶⁹	31	0.91	0.75		SF-36: -0.66 ; Pain VAS: 0.54 ($P < 0.001$)
Thai ⁵	63	0.62	0.93		
Turkish ⁷⁰	95	0.94	0.89–0.91		Pain VAS: 0.367 ($P < 0.01$), Schober test -0.068 (NS), RMDQ: 0.815 ($P < 0.001$)
Quebec Back Pain Disability Scale					
Dutch ⁷³	120	0.90	0.95		RMDQ:0.80, Pain VAS:0.70, Course of the complaint:0.35
French ⁷⁴	32				Dallas: 0.755, Perceived health status: 0.739, Impairment score: 0.449, Pain VAS: 0.448, HADS: 0.473
Persian ⁶⁹	31	0.86	0.92		SF-36: -0.62 , Pain VAS: 0.46 ($P < 0.001$)
Turkish ⁷⁵	83		0.93		Pain VAS: 0.63 ($P < 0.001$)

(Continued)

Table 3. Continued

Language	Sample Size	Test-Retest Reliability (ICC)	Internal Consistency (Cronbach's alpha)	Responsiveness	Construct Validity (Pearson correlation)
Roland Morris Disability Questionnaire					
Arabic ⁷⁷	62		0.94	$r = 0.83$ ($P = 0.000$)	Pain VAS: 0.33 ($P = 0.0001$), Schober test: 0.27 ($P = 0.0001$), General Function Score: 0.56 ($P = 0.0001$)
Cantonese ⁷⁸	112	0.94		ROC: 0.71–0.89	Global rating of change in overall condition at discharge: -0.22 ($P = 0.02$), Mean change score of pain: -0.44 ($P < 0.001$)
Mandarin ¹⁰	112	0.81	0.84		SF-36: -0.73 , Pain index: 0.32 ($P < 0.05$)
Danish ⁹	135		0.94		SF-36: -0.88 ($P < 0.001$), LBP rating scale: 0.89 ($P < 0.001$)
Dutch ⁷⁹	30	0.91			
French ¹¹	80	0.89			Pain VAS: 0.27 ($P = 0.018$)
German ⁸⁰	125		0.81		Pain VAS: 0.81 ($P = 0.0001$), Forward bending: 0.48 ($P = 0.0001$), Lateral bending: -0.47 ($P = 0.0001$)
Greek ⁶⁶	697		0.88		ODI: 0.729 ($P < 0.0005$)
Italian ⁸¹	70	0.92	0.82		Pain VAS: 0.79 ($P < 0.001$)
Japanese ⁶⁷	97	0.95	0.86		ODI: 0.785 ($P < 0.01$), JOA: -0.568 ($P < 0.01$)
Norwegian ⁶⁸	105	0.89	0.94		ODI: 0.73, Pain VAS: 0.32
Persian ⁶⁹	31	0.86	0.83		SF-36: -0.62 , Pain VAS: 0.36 ($P < 0.001$)
Portuguese ⁸²	30	0.94–0.95			Pain VAS: 0.79 ($P < 0.01$), Pain numerical scale (0–6): 0.80 ($P < 0.01$)
Spanish ⁸³	195	0.87	0.83–0.91		Pain VAS: 0.347 ($P = 0.0000$), ODI: 0.197 ($P = 0.0061$)
Swedish ⁸⁴	72	0.88			Perceived disability: 0.64 ($P < 0.001$), Pain severity: 0.54 ($P < 0.001$), Perceived life control: -0.32 ($P < 0.01$), General activity: -0.27 ($P < 0.05$)
Thai ⁸⁵	120		0.83		
Turkish ⁸⁶	81	0.79–0.86	0.85–0.89	Effect size: -0.53	Spinal mobility: -0.33 ($P < 0.01$) -0.11 (NS)
Scoliosis Research Society 22					
Japanese ^{88,101}	141		0.57		Radiologic measures: -0.33 ($P < 0.01$)
Spanish ^{89,90}	175	0.89	0.96		Radiologic Measures: -0.25 ($P < 0.001$)
Turkish ⁹¹	82	0.76–0.82	0.76–0.83		SF-36: 0.27–0.75 ($P < 0.0001$)
Aberdeen Low Back Pain Disability Scale					
Cantonese ⁷	473	0.94	0.85	Effect size: 0.59–0.81	RMDQ at baseline: 0.66, RMDQ at discharge 0.68
German ⁹⁵	158	0.95	0.86	SRM: 1.36–1.96	SF-36: -0.48 to -0.76 ($P < 0.0001$), FFbH-R: -0.76 ($P < 0.0001$)

SF-36 indicates Short Form Health Survey; BASFI, Bath Ankylosing Spondylitis Functional Index; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; Pain VAS, Pain Visual Analogue Scale; RMDQ, Roland Morris Disability Questionnaire; SEM, standard error of measurement; SRM, standardized response mean; WHOQOL-BREF, World Health Organization Quality of Life Assessment; ODI, Oswestry Disability Index; HADS, Hospital Anxiety and Depression Scale; FFbH-R, Hannover Functional Ability Questionnaire; NS, not significant.

where possible, the original name of the questionnaire should be retained and supplemented with the name of the language/country to which the questionnaire has been adapted.

This is the first study to systematically review self-report measures for LBP patients that are available in different languages. While there is a large number of original questionnaires available, very few have been translated into other languages, particularly languages commonly spoken like Mandarin (885 million speakers by 2 questionnaires), Spanish (332 million speakers by 8 questionnaires), Bengali (322 million speakers by no questionnaire), Hindi (182 million speakers by no questionnaire), Portuguese (181 million speakers by 1 questionnaire), Russian (145 million speakers by no questionnaire), Japanese (127 million speakers by 3 questionnaires), and German (120 million speakers by 8 questionnaires).¹⁶ This pattern of results suggests that there is no association between the number of people who speak a language and the number of adaptations to that language.

The translation procedures used were quite similar to those recommended to those in the *Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures*¹; however, the psychometric testing of the adapted questionnaire varied substantially between studies. There are major differences in sample sizes, test-

retest periods, statistical analysis, and benchmarks considering reliability and validity.

The Terwee quality criteria² do not summate into one overall quality score, as is sometimes done with trial methodologic quality scales. Summing of items presumes that the items assess the same attribute and that each item is equally important, a presumption that may not be true. Additionally, summed scores do not indicate the specific methodologic problems that are most prevalent. Accordingly, we think that it is more informative to separately consider each of the items of the Terwee scale. In this study, most of the adapted questionnaires were properly tested for reliability and construct validity, but not for internal consistency. Responsiveness and the assessment of floor and ceiling effects were usually not tested at all. We recommend, for further cross-cultural adaptations, that special attention be paid to assessment of internal consistency, responsiveness tests, and floor and ceiling effects.

One common problem that can be observed in our results is that some authors validated the self-report measures by correlating them with scales that do not measure the same construct. We hypothesize that the lack of validated questionnaires in some languages reinforces this procedure. For example, the Korean Oswestry¹⁷ was val-

Table 4. Quality Criteria for Psychometric Characteristics of LBP Questionnaires

Language	Internal Consistency	Construct Validity	Reliability	Responsiveness	Ceiling and Floor Effects
Acute Low Back Pain Screening Questionnaire					
Norwegian ²⁰	+	+	+	0	0
Bath Ankylosing Spondylitis Disease Activity Index					
Spanish ²⁸	0	+	+	?	0
Swedish ²⁹	0	+	0	?	0
Turkish ³⁰	?	+	—	?	0
Bath Ankylosing Spondylitis Functional Index					
German ¹⁴	?	+	+	?	0
Italian ¹⁵	?	+	+	?	0
Romanian ³²	?	—	+	0	0
Spanish ²⁸	0	+	—	?	0
Swedish ⁴²	0	+	?	?	0
Turkish ¹⁰⁰	?	+	+	0	0
Bournemouth Questionnaire					
Danish ³⁶	?	+	+	0	0
Dallas Pain Questionnaire					
French ⁴⁰	?	+	+	0	0
Core set					
Spanish ⁸	?	0	0	?	—
Dougados Functional Index for Ankylosing Spondylitis					
German ¹⁴	?	—	+	?	0
Italian ¹⁵	?	+	+	?	0
Spanish ²⁸	0	+	+	?	0
Turkish ⁴³	?	+	—	?	0
Fear Avoidance Beliefs Questionnaire					
French ⁴⁵	0	+	+	?	0
German ⁴⁶	+	+	+	0	0
Spanish ⁴⁷	?	+	+	0	—
Low Back Pain Rating Scale					
German ⁵⁵	+	+	?	0	0
NASS Lumbar Spine Outcome Assessment Instrument					
German ¹³	0	+	+	0	0
Italian ⁵⁸	?	+	+	0	0
Spanish ⁵⁹	+	+	+	+	—
Oswestry Disability Index					
Arabic ¹²	+	+	+	0	0
Mandarin ⁶⁴	+	0	+	0	0
Danish ⁶	?	+	+	0	+
German ⁶⁵	?	+	+	0	0
Greek ⁶⁶	?	+	0	0	0
Japanese ⁶⁷	?	+	+	0	0
Korean ¹⁷	?	+	+	0	0
Norwegian ⁶⁸	?	+	+	0	0
Persian ⁶⁹	?	+	+	0	0
Thai ⁵	?	0	—	0	0
Turkish ⁷⁰	?	—	+	0	0
Quebec Back Pain Disability Scale					
Dutch ⁷³	?	+	+	0	0
French ⁷⁴	?	+	0	0	0
Persian ⁶⁹	?	+	+	0	0
Turkish ⁷⁵	?	+	0	0	0
Roland Morris Disability Questionnaire					
Arabic ⁷⁷	?	+	0	?	0
Cantonese ⁷⁸	?	+	+	+	0
Mandarin ¹⁰	?	+	+	0	0
Danish ⁹	?	+	0	0	0
Dutch ⁷⁹	0	0	+	0	0
French ¹¹	0	+	+	0	0
German ⁸⁰	?	+	?	0	0
Greek ⁶⁶	?	+	0	0	0
Italian ⁸¹	?	+	+	0	0
Japanese ⁶⁷	?	+	+	0	0
Norwegian ⁶⁸	?	+	+	0	0
Persian ⁶⁹	?	+	+	0	0
Portuguese ⁸²	0	+	+	0	0
Spanish ⁸³	?	+	+	0	0
Swedish ⁸⁴	0	+	+	0	0
Thai ⁸⁵	?	0	0	0	0
Turkish ⁸⁶	0	—	+	?	0

(Continued)

Table 4. Continued

Language	Internal Consistency	Construct Validity	Reliability	Responsiveness	Ceiling and Floor Effects
Scoliosis Research Society 22					
Japanese ^{88,101}	?	+	0	0	0
Spanish ^{89,90}	+	+	+	0	—
Turkish ⁹¹	?	+	+	0	0
Aberdeen Low Back Pain Disability Scale					
Cantonese ⁷	?	+	+	?	0
German ⁹⁵	?	+	+	?	0

+ indicates positive rating; ?, indeterminate rating; —, negative rating; 0, no information available.

idated by comparison to pain and quality of life measures because at that time there was not another disability measure for LBP available.

This article clearly shows that there is an enormous lack of cross-cultural adaptations of LBP questionnaires. The adapted questionnaires varied considerably in the extent of psychometric testing each had undergone. Our study shows that there is a clear need for further work in this area.

Key Points

- There have been no systematic reviews of cross-cultural adaptations of LBP self-report outcome measures.
- Our review located 40 LBP self-report measures, of which only 15 had been adapted to a new language.
- There is an urgent need for further cross-cultural adaptation.



document

Appendix available online only through Article Plus.

Acknowledgments

The authors thank Raymond Ostelo, Kitir Jirarattana-phochochai, Roberto Padua, Steven Linton, Gro Jamtvedt, Kare Birger, Petros Boscaínos, Hanne Albert, Henrik Lauridsen, Annie Mannion, Dennis Turk, Paul Bettie, Bart Koes, Francisco Kovacs, Margreth Grotle, Rieckie de Vet, and Caroline Terwee for assisting us in the Results section.

References

1. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
2. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria are proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
3. Magnussen L, Strand LI, Lygren H. Reliability and validity of the Back Performance Scale: observing activity limitation in patients with back pain. *Spine* 2004;29:903–7.
4. Mannion AF, Elfering A, Staerle R, et al. Outcome assessment in low back pain: how low can we go? *Eur Spine J* 2005;14:1014–26.
5. Surachai SJ, Hunzavong T, Jirarattana-phochai J. Reliability of Thai version of Oswestry Questionnaire for the evaluation of low back pain patients. *Srinagarind Med J* 2002;17:248–51.
6. Lauridsen HH, Hartvigsen J, Manniche C, et al. Danish version of the Oswestry Disability Index for patients with low back pain: 1. Cross-cultural adaptation, reliability, and validity in two different populations. *Eur Spine J* 2006;15:1705–16.
7. Leung ASL, Lam T, Hedley AJ, et al. Use of a subjective health measure on Chinese low back pain patients in Hong Kong. *Spine* 1999;24:961–6.
8. Pellise F, Alvarez L, Escudero O, et al. Metric characteristics of the six-question 'core-set' in the evaluation of back pain. *Eur Spine J* 2003;12(suppl):12–3.
9. Albert HB, Jensen AM, Dahl D, et al. Criteria validation of the Roland Morris questionnaire: a Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica. *Ugeskr Læger* 2003;165:1875–80.
10. Chen SM, Liu MF, Wang BM, et al. Chinese translation and adaptation of the Roland-Morris Low Back Pain Disability Questionnaire. *FJPT* 2003;28:324–32.
11. Coste J, Le Parc JM, Berge E, et al. French validation of a disability rating scale for the evaluation of low back pain (EIFEL questionnaire). *Rev Rhum* 1993;60:335–41.
12. Guermazi M, Mezghani M, Ghroubi S, et al. The Oswestry index for low back pain translated into Arabic and validated in an Arab population. *Ann Readapt Med Physique* 2005;48:1–10.
13. Pose B, Sangha O, Peters A, et al. Validation of the North American Spine Society Instrument for assessment of health status in patients with chronic backache. *Z Orthop Ihre Grenzgeb* 1999;137:437–41.
14. Ruof J, Sangha O, Stucki G. Evaluation of a German version of the Bath Ankylosing Spondylitis Functional Index (BASFI) and Dougados Functional Index (DFI). *Z Rheumatol* 1999;58:218–25.
15. Salaffi F, Stancati A, Silvestri A, et al. Validazione delle versioni italiane del Bath Ankylosing Spondylitis Functional Index (BASFI) e del Dougados Functional Index (DFI) in pazienti con spondilite anchilosante. *Reumatismo* 2005;57:161–73.
16. Gordon RG. *Ethnologue: Languages of the World*, 15th ed. Dallas: SIL International, 2005.
17. Kim DY, Lee SH, Lee HY, et al. Validation of the Korean version of the Oswestry Disability Index. *Spine* 2005;30:E123–7.
18. Streiner DL, Normam GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 3rd ed. New York: Oxford University, 2003.
19. Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and sub acute back pain. *Clin J Pain* 1998;14:209–15.
20. Grotle M, Vollestad NK, Brox JI. Screening for yellow flags in first-time acute low back pain: reliability and validity of the Acute Low Back Pain Screening Questionnaire. *Clin J Pain* 2006;22:458–67.
21. Ettinger B, Black DM, Nevitt MC. Contribution of vertebral deformities to chronic back pain and disability. *J Bone Miner Res* 1992;7:449–56.
22. Symonds TL, Burton AK, Tillotson KM, et al. Absence resulting from low back trouble can be reduced by psychosocial intervention at the work place. *Spine* 1995;20:2738–45.
23. Tesio L, Granger CV, Fiedler RC. A unidimensional pain/disability measure for low back pain syndromes. *Pain* 1997;69:269–78.
24. Stratford PW, Binkley JM, Riddle DL. Development and initial validation of the Back Pain Functional Scale. *Spine* 2000;25:2095–102.
25. Rytokoski U, Puukka P, Talo S. Biodisability in relation to other disease consequences in the functional assessment of patients with chronic low back pain. *Int J Rehabil Res* 1997;20:225–44.
26. Davidson M, Keating JL, Eyres S. A low back-specific version of the SF-36 Physical Functioning Scale. *Spine* 2004;29:586–94.
27. Garret S, Jenkinson T, Kennedy LG, et al. A new approach to defining disease status in ankylosing spondylitis: the development of the Bath An-

- kylosing Spondylitis Disease Activity Index. *J Rheumatol* 1994;21:2286–91.
28. Cardiel MH, Londoño JD, Gutiérrez E, et al. Translation, cross-cultural adaptation, and validation of the Bath Ankylosing Spondylitis Functional Index (BASFI), the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the Dougados Functional Index (DFI) in a Spanish speaking population with spondyloarthropathies. *Clin Exp Rheumatol* 2003;21:451–8.
 29. Waldner A, Cronstedt H, Stenstrom CH. The Swedish version of the Bath Ankylosing Spondylitis Disease Activity Index: reliability and validity. *Scand J Rheumatol* 1999;28:10–6.
 30. Akkoc Y, Karatepe AG, Akar S, et al. A Turkish version of the Bath Ankylosing Spondylitis Disease Activity Index: reliability and validity. *Rheumatol Int* 2005;25:280–4.
 31. Calin A, Garrett S, Whitelock H. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol* 1994;21:2281–5.
 32. Udea G, Ciobanu C, Mihai C, et al. Evaluation of a Romanian version of the Bath Ankylosing Spondylitis Functional Index (BASFI) in patients with spondyloarthropathies. *Rom J Intern Med* 2004;42:199–209.
 33. Karatepe AG, Akkoc Y, Akar S, et al. The Turkish versions of the Bath Ankylosing Spondylitis and Dougados Functional Indices: reliability and validity. *Rheumatol Int* 2005;25:612–8.
 34. Yanik B, Gürsel YK, Kutlay S, et al. Adaptation of the Bath Ankylosing Spondylitis Functional Index to the Turkish population, its reliability and validity: functional assessment in AS. *Clin Rheumatol* 2005;24:41–7.
 35. Bolton JE, Breen AC. The Bournemouth Questionnaire: a short-form comprehensive outcome measure: I. Psychometric properties in back pain patients. *J Manipulative Physiol Ther* 1999;22:503–10.
 36. Hartigsen J, Lauridsen H, Ekstrom S, et al. Translation and validation of the Danish version of the Bournemouth questionnaire. *J Manipulative Physiol Ther* 2005;28:402–7.
 37. Deyo RA, Battie M, Beurskens AJ. Outcome measures for low back pain research: a proposal for standardized use. *Spine* 1998;23:2003–13.
 38. Harper AC, Harper DA, Lambert LJ, et al. Development and validation of the Curtin Back Screening Questionnaire (CBSQ): a discriminative disability measure. *Pain* 1995;60:73–81.
 39. Lawlis GF, Cuencas R, Selby D. The development of the Dallas Pain Questionnaire: an assessment of the impact of spinal pain on behaviour. *Spine* 1989;14:511–6.
 40. Martyr M, Blotman F, Avouac B, et al. Validation of the French version of the Dallas Pain Questionnaire in chronic low back pain patients. *Rev Rhum Engl Ed* 1998;65:126–34.
 41. Dougados M, Gueguen A, Nakache JP, et al. Evaluation of a functional index and an articular index in ankylosing spondylitis. *J Rheumatol* 1988;15:302–7.
 42. Cronstedt H, Waldner A, Stenstrom CH. The Swedish version of the Bath Ankylosing Spondylitis Functional Index: reliability and validity. *Scand J Rheumatol* 1999;28(suppl):1–8.
 43. Ozer HTE, Sarpel T, Gulek B, et al. Evaluation of the Turkish version of the Dougados functional index in ankylosing spondylitis. *Rheumatol Int* 2005;25:368–72.
 44. Waddell G, Newton M, Henderson I, et al. A Fear Avoidance Beliefs Questionnaire (FABQ) and the role of fear avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157–68.
 45. Chaory K, Fayad F, Rannou F, et al. Validation of the French version of the Fear Avoidance Belief Questionnaire. *Spine* 2004;29:908–13.
 46. Pflingsten M, Kroner-Herwig B, Leibing E, et al. Validation of the German version of the Fear-Avoidance Beliefs Questionnaire (FABQ). *Eur J Pain* 2000;4:259–66.
 47. Kovacs FM, Muriel A, Medina JM, et al. Psychometric characteristics of the Spanish version of the FAB Questionnaire. *Spine* 2006;31:104–10.
 48. Nork SE, Hu SS, Workman KL. Patient outcomes after decompression and instrumented posterior spinal fusion for degenerative spondylolisthesis. *Spine* 1999;24:561–9.
 49. Feise JF, Menke M. Functional Rating Index: a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine* 2001;26:78–87.
 50. Hagg O, Fritzel P, Romberg K. The General Function Score: a useful tool for measurement of physical disability: validity and reliability. *Eur Spine J* 2001;10:203–10.
 51. Kohlmann T, Raspe H. The Hannover Functional Ability Questionnaire for measuring low back pain-related limitations (FFbH-R). *Rehabilitation* 1996;35:I–VIII.
 52. Greenough CG, Fraser RD. Assessment of outcome in patients with low-back pain. *Spine* 1992;17:36–41.
 53. Holt AE, Shaw NJ, Shetty A, et al. The reliability of the Low Back Outcome for back pain. *Spine* 2002;27:206–10.
 54. Manniche C, Asmussen K, Lauritsen B. Low Back Pain Rating Scale: validation of a tool for assessment of low back pain. *Pain* 1994;57:317–26.
 55. Nuhr MJ, Crevenna R, Quittan M, et al. Cross-cultural adaptation of the Manniche questionnaire for German-speaking low back pain patients. *J Rehabil Med* 2004;36:267–72.
 56. Walsh TL, Hanscom B, Lurie JD, et al. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine* 2003;28:607–15.
 57. Daltroy LH, Cats-Baril WL, Katz JN, et al. The North American Spine Society Lumbar Spine Outcome Assessment Instrument: reliability and validity tests. *Spine* 1996;21:741–9.
 58. Padua R, Padua L, Ceccarelli E, et al. Cross-cultural adaptation of the Lumbar North American Spine Society Questionnaire for Italian-speaking patients with lumbar spinal disease. *Spine* 2001;26:344–7.
 59. Sarasqueta C, Gabaldon O, Iza I, et al. Cross-cultural adaptation and validation of the NASS outcomes instrument in Spanish patients with low back pain. *Eur Spine J* 2005;14:586–94.
 60. Kopec KA, Esdaile JM. Occupational role performance in persons with back pain. *Disabil Rehabil* 1998;20:373–9.
 61. Stucki G, Daltroy LH, Liang MH. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine* 1996;21:796–803.
 62. Pratt RK, Fairbank JCT, Virr A. The reliability of the Shuttle Walking Test, the Swiss Spinal Stenosis Questionnaire, the Oxford Spinal Stenosis Score, and the Oswestry Disability Index in the assessment of patients with lumbar spinal stenosis. *Spine* 2002;27:84–91.
 63. Fairbank JCT, Couper J, Davies JB. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
 64. Chow JH, Chan CC. Validation of the Chinese version of the Oswestry Disability Index. *Work* 2005;25:307–14.
 65. Mannion AF, Junge A, Fairbank JCT, et al. Development of a German version of the Oswestry Disability Index: 1. Cross-cultural adaptation, reliability, and validity. *Eur Spine J* 2006;15:5–65.
 66. Boscaino PJ, Sapkas G, Stilianesi E, et al. Greek versions of the Oswestry and Roland-Morris Disability Questionnaires. *Clin Orthop* 2003;411:40–53.
 67. Fujiwara A, Kobayashi N, Saiki K, et al. Association of the Japanese Orthopaedic Association score with the Oswestry Disability Index, Roland Morris Questionnaire and short form 36. *Spine* 2003;28:1601–7.
 68. Grotle M, Brox JI, Vollestad NK. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. *J Rehabil Med* 2003;35:214–47.
 69. Mousavi SJ, Parnianpour M, Mehdian H, et al. The Oswestry Disability Index, the Roland Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: translation and validation studies of the Iranian versions. *Spine* 2006;31:E454–9.
 70. Yakut E, Duger T, Oksuz C, et al. Validation of the Turkish version of the Oswestry Disability Index for patients with low back pain. *Spine* 2004;29:581–5.
 71. Gallon RL. Perception of disability in chronic back pain patients: a long term follow up. *Pain* 1989;37:67–75.
 72. Kopec KA, Esdaile JM, Abrahamowicz M. The Quebec Back Pain Disability Scale: conceptualization and development. *J Clin Epidemiol* 1996;49:151–61.
 73. Schoppink LE, van Tulder MW, Koes BW, et al. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. *Phys Ther* 1996;76:268–75.
 74. Yvanes-Thomas M, Calmels P, Bethoux F, et al. Validity of the French-language version of the Quebec back pain disability scale in low back pain patients in France. *Joint Bone Spine* 2002;69:397–405.
 75. Bicer A, Yazici A, Camdeviren H, et al. Assessment of pain and disability in patients with chronic low back pain: reliability and construct validity of the Turkish version of the Quebec Back Pain Disability Scale and Pain Disability Index. *J Back Musculoskeletal Rehabil* 2005;18:37–44.
 76. Roland M, Morris R. A study of the natural history of back pain: I. Development of a reliable and sensitive measure of disability of low back pain. *Spine* 1983;8:141–4.
 77. Bejia I, Younes M, Kamel BS, et al. Validation of the Tunisian version of the Roland-Morris questionnaire. *Eur Spine J* 2005;14:171–4.
 78. Tsang RCC. Measurement properties of the Hong Kong Chinese version of the Roland-Morris Disability Questionnaire. *Hong Kong Physiother J* 2004;22:40–9.
 79. Brouwer S, Kuijer W, Dijkstra P, et al. Reliability and stability of the Roland Morris Disability Questionnaire. *Disabil Rehabil* 2004;26:162–5.
 80. Wiesinger GF, Nuhr M, Quittan M, et al. Cross-cultural adaptation of the Roland Morris questionnaire for German-speaking patients with low back pain. *Spine* 1999;24:1099–103.
 81. Padua R, Padua L, Ceccarelli E, et al. Italian version of the Roland Disabil-

- ity Questionnaire, specific for low back pain: cross-cultural adaptation and validation. *Eur Spine J* 2002;11:126–9.
82. Nusbaum L, Natour J, Ferraz MB, et al. Translation, adaptation and validation of the Roland Morris questionnaire-Brazil Roland Morris. *Braz J Med Biol Res* 2001;34:203–10.
 83. Kovacs FM, Llobera J, Gil Del Real MT, et al. Validation of the Spanish version of the Roland Morris questionnaire. *Spine* 2002;27:538–42.
 84. Johansson E, Lindberg P. Sub acute and chronic low back pain. reliability and validity of a Swedish version of the Roland and Morris Disability Questionnaire. *Scand J Rehabil Med* 1998;30:139–43.
 85. Jirarattanaphochai J, Jung S, Sumananont C, et al. Reliability of the Roland-Morris disability questionnaire (Thai version) for the evaluation of low back pain patients. *J Med Assoc Thai* 2005;88:407–11.
 86. Kucukdeveci AA, Tennant A, Elhan AH, et al. Validation of the Turkish version of the Roland Morris Disability Questionnaire for use in low back pain. *Spine* 2001;26:2738–43.
 87. Climent JM, Reig A, Sanchez J. Construction and validation of a specific quality of life instrument for adolescents with spine deformities. *Spine* 1995;20:2006–11.
 88. Watanabe K, Hasegawa K, Hirano T, et al. Use of the Scoliosis Research Society Outcomes Instrument to evaluate patient outcome in untreated idiopathic scoliosis patients in Japan: 1. Comparison with non scoliosis group: preliminary /limited review in a Japanese Population. *Spine* 2005; 30:1197–201.
 89. Climent JM, Bago R, Ey A, et al. Validity of the Spanish version of the Scoliosis Research Society-22 (SRS-22) Patient Questionnaire. *Spine* 2005; 30:705–9.
 90. Bago J, Climent JM, Ey A, et al. The Spanish Version of the SRS-22 Patient Questionnaire for idiopathic scoliosis: transcultural adaptation and reliability analysis. *Spine* 2004;29:1676–80.
 91. Alanay A, Cil A, Berk H, et al. Reliability and validity of adapted Turkish version of the Scoliosis Research Society-22 (SRS-22) Questionnaire. *Spine* 2005;30:2464–8.
 92. Robinson RC, Kishino N, Matheson L, et al. Improvement in postoperative and nonoperative spinal patients on a self-report measure of disability: the Spinal Function Sort (SFS). *J Occup Rehabil* 2003;13:107–13.
 93. Itzkovich M, Catz A, Tamir A. Spinal pain independence measure: a new scale for assessment of primary ADL dysfunction related to low back pain. *Disabil Rehabil* 2001;23:186–91.
 94. Ruta DA, Garrat AM, Wardlaw D, et al. Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 1994;19: 1887–96.
 95. Osthus H, Cziske R, Jacobi E. A German version of the Extended Aberdeen Back Pain Scale. *Spine* 2006;31:571–7.
 96. Atlas SJ, Deyo RA, van den Ancker M, et al. The Maine-Seattle Back Questionnaire: a 12-item disability questionnaire for evaluating patients with lumbar sciatica or stenosis. Results of a derivation and validation cohort analysis. *Spine* 2003;28:1869–76.
 97. Hazard RG, Haugh LD, Reid S, et al. Early prediction of chronic disability after occupational low back injury. *Spine* 1996;21:945–51.
 98. Waddell G, Main CJ. Assessment of severity in low back pain disorders. *Spine* 1984;9:204–8.
 99. Sanders JO, Polly DW Jr, Cats-Baril WL, et al. Analysis of patient and parent assessment of deformity in idiopathic scoliosis using the Walter Reed Visual Assessment Scale. *Spine* 2003;18:2158–63.
 100. Ozer HTE, Sarpel T, Gulek B, et al. The Turkish version of the Bath Ankylosing Spondylitis Functional Index: reliability and validity. *Clin Rheumatol* 2005;24:123–8.
 101. Watanabe K, Hasegawa K, Hirano T, et al. Use of the Scoliosis Research Society Outcomes Instrument to evaluate patient outcome in untreated idiopathic scoliosis patients in Japan: 2. Relation between spinal deformity and patient outcomes. *Spine* 2005;30:1202–5.

Appendix 1: Quality criteria for measurement properties of health status questionnaires²

Quality criteria for measurement properties of health status questionnaires

Property	Quality criteria
Internal consistency	<ul style="list-style-type: none"> + factor analyses performed on adequate sample size ($7 * \# \text{ items}$ and ≥ 100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70-0.95; ? no factor analysis OR doubtful design or method; - Cronbach's alpha(s) < 0.70 or > 0.95, despite adequate design and method; 0 no information found on internal consistency.
Construct validity	<ul style="list-style-type: none"> + specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses; ? doubtful design or method (e.g. no hypotheses); - less than 75% of hypotheses were confirmed, despite adequate design and methods; 0 no information found on construct validity.
Reliability	<ul style="list-style-type: none"> + ICC or Kappa ≥ 0.70; ? doubtful design or method (e.g. time interval not mentioned); - ICC or Kappa < 0.70, despite adequate design and method; 0 no information found on reliability.
Responsiveness	<ul style="list-style-type: none"> + Smallest detectable change_{individual} or Smallest detectable change_{group} $<$ Minimal important change OR Minimal important change outside the limits of agreement OR Responsiveness ratio > 1.96 OR Area under the curve ≥ 0.70; ? doubtful design or method OR sample size < 50 OR methodological flaws; - Smallest detectable change_{individual} or Smallest detectable change_{group} \geq Minimal important change OR Minimal important change equals or inside limits of agreement OR Responsiveness ratio ≤ 1.96 OR Area under the curve < 0.70, despite adequate design and methods; 0 no information found on responsiveness.
Floor and ceiling effects	<ul style="list-style-type: none"> + $\leq 15\%$ of the respondents achieved the highest or lowest possible scores; ? doubtful design or method OR sample size < 50 OR methodological flaws; - $> 15\%$ of the respondents achieved the highest or lowest possible scores, despite adequate design and methods; 0 no information found on interpretation.

+ = positive rating; ? = indeterminate rating; - = negative rating; 0 = no information available

Chapter Eight

Psychometric characteristics of the Brazilian-Portuguese versions of the Functional Rating Index and the Roland Morris Disability Questionnaire

Chapter Eight is published as:


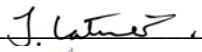
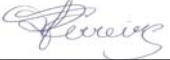

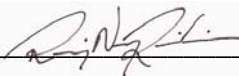
Costa LOP, Maher CG, Latimer J, Ferreira PH, Pozzi GC, Ribeiro RN (2007)

Psychometric characteristics of the Brazilian-Portuguese versions of the Functional Rating Index and the Roland Morris Disability Questionnaire. *Spine* 32:1902-190.

**Statement from co-authors confirming authorship contribution of the
PhD candidate**

As co-authors of the paper “Psychometric characteristics of the Brazilian-Portuguese versions of the Functional Rating Index and the Roland Morris Disability Questionnaire”, we confirm that Leonardo Costa has made the following contributions:

- Conception and design of the research
- Data collection
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

Christopher G Maher		Date	<u>20.03.2009</u>
Jane Latimer		Date	<u>20.03.2009</u>
Paulo H Ferreira		Date	<u>20.03.2009</u>
Giovanni C Pozzi		Date	<u>20.03.2009</u>
Rodrigo N Ribeiro		Date	<u>20.03.2009</u>

Psychometric Characteristics of the Brazilian-Portuguese Versions of the Functional Rating Index and the Roland Morris Disability Questionnaire

Leonardo Oliveira Pena Costa, PT, MSc,*† Chris G. Maher, PT, PhD,* Jane Latimer, PT, PhD,*
Paulo Henrique Ferreira, PT, PhD,*‡ Giovanni Campos Pozzi, PT,† and Rodrigo Nogueira Ribeiro, PT†

Study Design. Translation, cross-cultural adaptation and psychometric testing of self-report outcome measurements.

Objectives. The aims of this study were to adapt the Functional Rating Index (FRI) to Brazilian-Portuguese and to test the psychometric properties of this new questionnaire and the Brazilian-Roland Morris Disability Questionnaire (RMDQ).

Summary of Background Data. Self-report measures are commonly used in clinical practice and in research studies. Most existing questionnaires were developed in English and there is only 1 Brazilian-Portuguese cross-cultural adaptation of a spine outcome measure.

Methods. The FRI was translated and adapted into Brazilian-Portuguese according to the *Guidelines for the process of cross-cultural adaptation of self report measures*. The Brazilian-Portuguese FRI and RMDQ were tested for internal consistency, reliability, ceiling and floor effects, construct validity, and responsiveness in 140 low back pain (LBP) patients.

Results. A very high level of internal consistency (Cronbach's $\alpha = 0.92$) and reliability [ICC = 0.95 (95% confidence interval, 0.93–0.97)] was shown for both instruments. The FRI and RMDQ were highly correlated ($r = 0.80$), while both the FRI and RMDQ were moderately correlated with pain at baseline ($r = 0.67$ and 0.55 , respectively). No ceiling or floor effects were detected; however, the responsiveness of both questionnaires was quite small (RMDQ ES = 0.10 [84% confidence interval, 0.04–0.16] and FRI ES = 0.18 [84% confidence interval, 0.11–0.24]).

Conclusion. The results of this study indicate that the Brazilian-Portuguese versions of the FRI and RMDQ are reliable and valid instruments for the measurement of disability in Brazilian-Portuguese-speaking patients with LBP presenting for physiotherapy treatment. Both instruments are suitable for use in clinical practice and research studies.

Key words: low back pain, functional status, Roland Morris Disability Questionnaire, Functional Rating Index, Portuguese version, reliability, construct validity. **Spine** 2007;32:1902–1907

Assessment of functional status by self-report questionnaires is common in clinical practice and research. A large number of generic and disease-specific instruments have been developed to assess the functional status of back pain patients; however, the majority have been developed for English speakers. This is a major problem because most of the world does not speak English. There are several ways to redress this problem. Researchers in non-English-speaking countries could develop their own self-report measures or adaptations of existing English language questionnaires could be performed, this latter method being simpler and more efficient. Cross-cultural adaptation of existing English language questionnaires would enable comparisons of different populations and permit the exchange of information across cultural and linguistic barriers.

Portuguese is 1 of the top 10 most widely spoken languages in the world and is used in Europe, Africa, and South America. It is the language spoken in Brazil, a country with a population of more than 180 million. A recent systematic review of cross-cultural adaptations of self-report measures for back pain patients¹ found that only the Roland Morris Disability Questionnaire (RMDQ) had been translated to Brazilian-Portuguese.² Limited psychometric testing has been performed on the Brazilian-Portuguese version using a small sample of 30 low back pain (LBP) subjects. Results demonstrate high test-retest reliability (intraclass correlation coefficient [ICC] = 0.95). Construct validity was tested by evaluating the correlation of the RMDQ with a visual analogue pain scale ($r = 0.79$) and with a 6-point pain numerical rating scale ($r = 0.80$). These results suggest that the RMDQ is moderately correlated with these pain measures.

The Functional Rating Index (FRI)³ is shorter and quicker to complete than the RMDQ and was developed to measure perceived disability in patients with back and/or neck pain. The FRI emphasizes function while concurrently measuring the patient's opinion, attitude, and self-rating of disability. The only cross-cultural adaptation of this instrument is a Turkish translation⁴ of the questionnaire, there being no Brazilian-Portuguese version. As this questionnaire can be used in patients

From the *Back Pain Research Group, School of Physiotherapy, University of Sydney, Lidcombe NSW, Australia; † Pontifícia Universidade Católica de Minas Gerais, Brazil; and ‡ Universidade Federal de Minas Gerais, Brazil.

Acknowledgment date: October 19, 2006. First revision date: January 14, 2007. Acceptance date: January 15, 2007.

Leonardo Costa is a PhD student supported by CAPES – Ministério da Educação – Brazil; Christopher Maher holds a research fellowship funded by the National Health and Medical Research Council of Australia.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint request to Leonardo Oliveira Pena Costa, PT, MSc, Back Pain Research Group, School of Physiotherapy, University of Sydney, East Street Lidcombe NSW Australia 2141; E-mail: lcos3060@usyd.edu.au

with both back and neck pain, it makes good sense to adapt the FRI to Brazilian-Portuguese.

The development of the FRI was based on the Oswestry Disability Index (ODI)⁵ and the Neck Disability Index.⁶ It is a 10-item questionnaire (8 items referring to activities of daily living that might be adversely affected by a spinal condition, and 2 items referring to different attributes of pain) with a 5-point Likert frequency scale. The patient ranks his or her perceived ability at the present time by selecting 1 of the 5 points of the scale (0 = no pain or full ability to function; 4 = worst possible pain and/or unable to perform this function at all). The total score is obtained by summing the item scores, dividing by the maximum possible total score (40) and multiplying by 100%. The range of scores is zero percent (no disability) to 100% (severe disability). The higher the score, the higher the perceived disability.

The RMDQ is a self-report questionnaire consisting of 24 items related to normal activities of daily living. The RMDQ was developed by selecting 24 relevant items from the longer Sickness Impact Profile.⁷ The questionnaire was transformed to a condition-specific measure for spinal disorders by adding the phrase “because of my back” to each statement. The patients are asked to circle those items, which they perceive as difficult to perform due to back pain. Each answer is scaled simply 0 or 1, thus leaving a range of scores of 0 to 24, a higher score indicating higher disability. The RMDQ has already been translated and adapted into 17 different languages, including Brazilian-Portuguese.^{2,8–23}

The English language FRI has had its psychometric properties tested (reliability, validity, responsiveness, clinical utility, and effect of socio-demographics) with excellent results.³ Another study²⁴ has compared the responsiveness and reliability of the FRI and the 18-item RMDQ, this study showing similar reliability between the questionnaires. However, the responsiveness was higher in the FRI (RMDQ effect size (ES): 0.44 and FRI ES: 0.64). Additionally, the minimum detectable change was lower in the FRI, suggesting that the FRI is better than the RMDQ in detecting true change in patients. It would therefore be well suited for use in clinical trials as a more responsive questionnaire requires a smaller sample size than a less responsive one. The authors concluded that “the FRI seems preferable to the 18-item RMDQ for using in clinical trials and clinical practice.”²⁴

These results suggest that it would be worthwhile developing a Brazilian-Portuguese version of the FRI and evaluating the psychometric properties of this. In order to evaluate construct validity, a comparison with the Brazilian-Portuguese RMDQ would be useful, in addition to evaluating the responsiveness of both the Brazilian-Portuguese FRI and RMDQ. The findings of this research will help Brazilian-Portuguese-speaking clinicians and researchers to determine which questionnaire is most suitable to their needs.

■ Methods

Overview of Study Design. The study was carried out in 2 phases: the first phase was the translation and cross-cultural adaptation of the FRI; the second phase was a study testing the psychometric properties of the Brazilian-Portuguese FRI and RMDQ.

Translation and Cross-Cultural Adaptation. The cross-cultural adaptation was performed in accordance with the Guidelines for the process of cross-cultural adaptation of self report measure²⁵ as follows:

1. Initial translation: Two independent translators translated the FRI from English to Brazilian-Portuguese.
2. Synthesis of the translations: After discussion, the 2 translators produced a consensus version of the FRI.
3. Back translation: Two new translators without a health sciences background independently back-translated the consensus version of the Brazilian-Portuguese FRI into English.
4. An expert committee reviewed all translations and discussed with the original translators possible discrepancies, and developed the final FRI version to be tested in Brazil.

Testing Psychometric Properties. This phase was performed on patients recruited from different physiotherapy clinics (public and private) in Belo Horizonte, Brazil, with the following inclusion criteria: presence of nonspecific acute or chronic low back pain, able to speak and read Portuguese, and aged between 18 and 80 years.

Determination of our sample size was based on advice contained in the *Quality criteria for health status questionnaires*.²⁶ The authors of this paper advise that at least 100 patients are necessary for internal consistency analysis and 50 for appropriate analysis of other tests (construct validity, reliability, and ceiling/floor effects analysis). Because of the potential for loss to follow-up and missing data, we decided to recruit 140 subjects.

The patients completed the FRI, RMDQ, and the pain measure at baseline (on presentation to the clinic), 24 hours later (for test-retest reliability), and following 4 weeks of treatment (for responsiveness).

Analysis. Further tests were conducted on the psychometric properties of the adapted questionnaires. A list of the tests follows:

1. Internal consistency (homogeneity) was evaluated using Cronbach's α and alpha “if item deleted”; Cronbach's α is considered adequate between 0.70 and 0.95.²⁶
2. Test-retest reliability (repeatability) was evaluated using the ICC(2,1); ICC has been interpreted as follows: <0.40, poor reliability; 0.40 to 0.75, moderate reliability; 0.75 to 0.90, substantial reliability; and >0.90, excellent reliability.²⁶
3. Construct validity was evaluated by correlating the Brazilian-Portuguese RMDQ, the Brazilian-Portuguese FRI, and an 11-point Pain Numerical Rating Scale (Pain NRS) at baseline using Pearson's r ; a score of 0.70 was recommended for instruments that measure the same construct. When similar constructs are compared, scores lower than 0.70 should be accepted.²⁶
4. Potential ceiling and floor effects were measured by assessing the distribution of answers across categories calculating the percentage of patients indicating the minimum and

maximum possible scores in both questionnaires. Ceiling and floor effects are considered to be present if more than 15% of respondents achieved the lowest or highest possible total score (ceiling and floor effects are not related to individual items).²⁶

5. Internal responsiveness of the Brazilian-Portuguese FRI and Brazilian-Portuguese RMDQ was assessed by calculating the Effect Size (ES)²⁴ with 84% confidence interval (CI). We calculated 84% CIs for direct comparison of the ES. We chose 84% CIs because nonoverlapping 84% CIs are equivalent to a Z test of means at the 0.05 level.²⁶ Responsiveness indicates the sensitivity of a questionnaire to measure true change. Higher scores are preferred.

The data were checked by 2 authors (G.C.P. and R.N.R.) and a random sample (10%) was double-checked by another author (L.O.P.C.) before the analysis.

■ Results

A total of 140 eligible patients were recruited and 138 answered all questionnaires at the 3 time points. Table 1 shows the characteristics of the study participants. The mean total scores for the FRI and RMDQ were 16.88 (SD 9.04) and 9.94 (SD 6.61), respectively, and the distribution of the total scores in both questionnaires at baseline are presented on Figure 1.

Both questionnaires achieved the same Cronbach's α value of 0.92. Analyzing alpha "if item deleted" demonstrated that no individual items from the RM, or FRI, contribute more to the construct than the others [FRI and RMDQ "if α deleted" (0.91–0.92)]. The FRI and the RMDQ had identical test-retest reliability, with an ICC (2,1) of 0.95 (95% CI, 0.93–0.97).

A high correlation between the FRI and RMDQ was found, a Pearson's $r = 0.80$ ($P < 0.001$) being obtained, and both the FRI and RMDQ are correlated moderately with Pain NRS at baseline [$r = 0.67$ ($P < 0.001$) and 0.55 ($P < 0.001$) respectively].

The ceiling and floor analysis was performed on the total score; the percentage of respondents who achieved the lowest score (floor effect) was 0.7% (1 patient) for the FRI and 1.4% (2 patients) for the RMDQ, none scored the highest score (ceiling effect) in both instruments. There were no floor and ceiling effects for the FRI and RMDQ total scores.

The RMDQ and FRI had similar levels of responsiveness; RMDQ ES = 0.10 (84% CI, 0.04–0.16) and FRI ES = 0.18 (84% CI, 0.11–0.24).

■ Discussion

Our study aimed to develop and test the psychometric properties of the Brazilian-Portuguese version of the FRI and to test the psychometric properties of the Brazilian-Portuguese RMDQ in LBP patients. The results of this study indicate that these versions of the FRI and RMDQ are reliable and valid instruments for the measurement of disability in Portuguese-speaking patients with LBP, making them suitable for use in routine health care and research studies. To provide benchmarks for our results, we calcu-

Table 1. Characteristics of Study Participants

Variable	All Study Participants*
Gender	
Male	65 (46.4)
Female	75 (53.6)
Age (yr)	41 (12)
LBP duration (wk)	154 (259)
Weight (kg)	70.1 (14.5)
Height (m)	1.68 (0.9)
Pain numerical rating intensity scale (0–10) at baseline	3.6 (1.4)
Pain interference (1–4) at baseline	2.8 (1.2)
Qualification level	
Elementary school (complete or incomplete)	37 (26.4)
School certificate	15 (10.7)
High school certificate	10 (7.1)
Diploma	36 (25.7)
Currently at university	23 (26.4)
Bachelor degree	10 (7.1)
Postgraduate certificate	9 (6.4)
Marital status	
Single	46 (32.9)
Married	71 (50.7)
Divorced	17 (12.1)
Widow	6 (4.3)
Working situation	
Yes	83 (59.3)
No	57 (40.7)
Previous back surgery	
Yes	6 (4.3)
No	134 (95.7)
Fitness level	
Active	65 (46.4)
Sedentary	75 (53.6)
Self-report health status	
Excellent	17 (12.1)
Very good	46 (32.9)
Good	39 (27.9)
Fair	30 (21.4)
Poor	8 (5.7)
Religion	
Catholic	101 (72.1)
Presbyterian	22 (15.7)
Buddhist	1 (0.7)
Other	16 (10.5)

*Continuous data are mean (SD); categorical data are N (%).

lated the mean reliability from the estimates of reliability reported in previous studies that evaluated the original English and non-English RMDQ (Table 2) and FRI (Table 3). This analysis confirmed that the Brazilian RMDQ, other non-English RMDQ and the original English-language RMDQ all have excellent test-retest reliability. It is noteworthy that the questionnaire performs equally well across a wide range of language/cultural settings.

The reliability of both instruments was tested by evaluating internal consistency and test-retest reliability. The high level of internal consistency achieved by both instruments confirms the homogeneity of the items; that is, the items are measuring a similar construct. Both the RMDQ and FRI satisfy the internal consistency benchmarks proposed recently for health status questionnaires.²⁶ Our results and those from other studies fall within the benchmark range of 0.70 to 0.95, suggesting an acceptable Cronbach's α . Similarly, the test-retest reliability achieved in this study and most others evaluat-

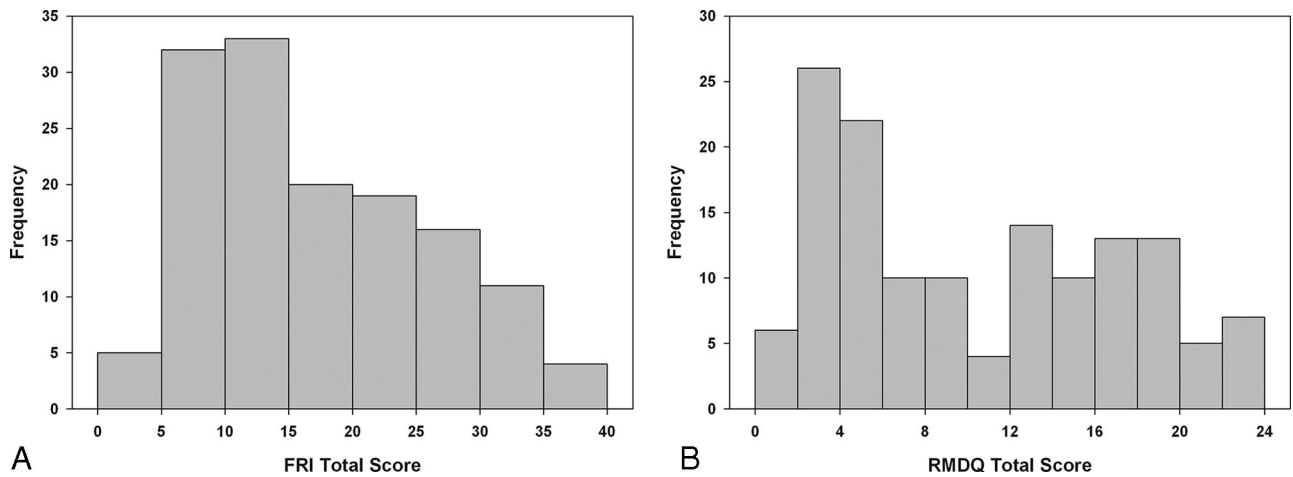


Figure 1. Distribution of the RMDQ and FRI total scores at baseline.

ing the FRI and the RMDQ meet or exceed the minimum standard for reliability²⁶ (ICC = 0.70). Our test-retest evaluation was performed with a 24-hour interval; therefore, a “memory effect” may have inflated these values; however, a longer interval between tests may also intro-

duce bias due to improvement in a patient’s condition following treatment.

The high intercorrelations observed between the FRI, RMDQ, and Pain NRS provide support for the construct validity of the FRI. The higher correlation between the FRI and Pain NRS could be explained as items 1 (pain intensity) and 7 (pain frequency) from the FRI are related to pain (20% of all items), while the RMDQ has only 1 pain-related item (4.2% of all items) (item 13 “my back is painful almost all the time”). The high correlation between the FRI and RMDQ showed a clear association between the 2 measures. To our knowledge, currently there is no “gold standard” for determining disability status of a patient with LBP; therefore, criterion validity was not able to be assessed.

The internal responsiveness of these instruments was tested with a test-retest design, with patients being treated in the period between tests. As such, it was expected that a change in scores was likely to occur due to the natural history of LBP, the effects of the physiotherapy intervention on the patient’s condition and/or the placebo effect. Surprisingly, patients did not appear to get better following intervention with the majority remaining unchanged. Responsiveness can be limited due to floor or ceiling effects; however, both tested instruments were not affected by ceiling or floor effects and, as

Table 2. Summary of Evidence on Reliability of English Language and Non-English Language Versions of the RMDQ

RMDQ Version	Sample Size	Test-Retest Reliability (ICC)	Internal Consistency (Cronbach’s α)
Non-English versions¹			
Arabic ⁸	62		0.94
Cantonese ⁹	112	0.94	
Mandarin ¹⁰	112	0.81	0.84
Danish ¹¹	135		0.94
Dutch ¹²	30	0.91	
French ¹³	80	0.89	
German ¹⁴	125		0.81
Greek ¹⁵	697		0.88
Italian ¹⁶	70	0.92	0.82
Japanese ¹⁷	97	0.95	0.86
Norwegian ¹⁸	105	0.89	0.94
Persian ²³	31	0.86	0.83
Portuguese ²	30	0.94	
This study	140	0.95	0.92
Spanish ¹⁹	195	0.87	0.83
Swedish ²⁰	72	0.88	
Thai ²¹	120		0.83
Turkish ²²	81	0.79	0.85
Mean	85	0.89*	0.86
English versions†			
Roland and Morris ²⁷	20	r = 0.91	
Deyo ²⁸	10	r = 0.83	
Kopec <i>et al</i> ²⁹	98	0.90	0.95
Stratford <i>et al</i> ³⁰	36	0.86	
Stratford and Binkley ³¹	136	0.79	0.87
Patrick <i>et al</i> ³²	52	0.76	0.90
Underwood <i>et al</i> ³³	48	0.89	0.96
Jacob <i>et al</i> ³⁴	96	0.93	0.88
Jensen <i>et al</i> ³⁵	50	r = 0.72	
Stratford <i>et al</i> ³⁶	28	0.81	0.87
Davidson and Keating ³⁷	47	0.53	
Mean	56	0.84*	0.91
Mean (all studies)	71	0.87*	0.88

*The mean of reliability tests across different studies was calculated by transforming the correlations into Z scores and after calculating the mean, the Z score was re-converted to a correlation.

†Some studies listed here used different versions of the RMDQ including the 18-, 23-, and 24-item questionnaires. All of the non-English versions are adaptations of the 24-Item RMDQ.

Table 3. Summary of Evidence on Reliability of English Language and Non-English Language Versions of the FRI

FRI Version	Sample Size	Test-Retest Reliability (ICC)	Internal Consistency (Cronbach’s α)
Non-English versions			
Turkish	84	0.92	0.96
English versions			
Chansirinukor <i>et al</i> ²⁴	143	0.67	
Bayar <i>et al</i> ³⁸	76	0.91	0.92
Childs and Piva ³⁹	131	0.63	
Mean (all studies)	108.5	0.83*	0.94
This study	140	0.95	0.92

*The mean of reliability tests across different studies was calculated by transforming the correlations into Z scores and after calculating the mean, the Z score was re-converted to a correlation.

a consequence, patients with low or high scores can be distinguished from each other. One possible reason for this poor responsiveness may be the chronicity of the LBP sample (LBP duration was 154 ± 259 weeks), which may have led to a poor outcome. Alternatively, patients may have changed, but the questionnaires were insufficiently sensitive to detect this change. Sixty of 138 patients (43.5%) scored less than 6 of 24 items on the RMDQ and 9 of 40 items on FRI at baseline, indicating that they had little disability and therefore the chances of these scores changing significantly was reduced. The only previous study to evaluate the responsiveness of the RMDQ and report the results in languages other than English was the Turkish version.²² This study found the RMDQ to have poor responsiveness, a small effect size of -0.53 being reported. A further responsiveness study comparing the English versions of the RMDQ (18-item) and the FRI found only small effect sizes as well (0.64 for FRI and 0.44 for RMDQ).²⁴ Ideally, responsiveness should be tested in a randomized clinical trial design or follow-up study using prevalidated instruments and a global assessment scheme as an external criterion of change.³ In this manner, internal and external responsiveness can be tested further.

We have tried to maximize the generalizability of our findings by using a broad inclusion criteria to select our samples (only illiterate patients were excluded), by choosing patients presenting for treatment of their spinal pain and by recruiting from different clinical settings (private and public physiotherapy clinics) across the entire city. There is no evidence that these patients are likely to be different from other patients presenting to primary care practitioners, such as general medical practitioners, chiropractors, and osteopaths for treatment of their spinal pain; we think that it is unlikely that the values yielded in this study will change dramatically in different samples for internal consistency, test-retest reliability, and construct validity. In samples with different levels of disability compared with this study, it is possible to find different results for responsiveness and ceiling and floor effects. Future studies should clarify these issues.

The results of our study demonstrate that the FRI and RMDQ are reliable and valid tools in the assessment of disability in Brazilian-Portuguese speaking patients with LBP presenting for physiotherapy treatment. It is now possible to perform intercultural comparisons between randomized clinical trials performed in Brazil and those performed in English-speaking countries.

■ Key Points

- A Brazilian-Portuguese version of the Functional Rating Index (FRI) was developed according to guidelines for cross-cultural adaptation and compared with an existing Brazilian-Portuguese Roland Morris Disability Questionnaire (RMDQ).

- Similar to the original English-language versions, the Brazilian-Portuguese FRI and RMDQ were shown to be valid and reliable.
- The 2 disability measures are useful for measuring LBP disability status and outcome in Brazilian-Portuguese speaking patients, allowing for intercultural comparisons.



Appendix available online through Article Plus.

Acknowledgments

The authors thank Marcos Amaral de Noronha, Gustavo Braz, and Daniela Faria e Lailah de Noronha for their assistance with the translation procedures.

References

1. Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. *Spine* 2007;32:1028–37.
2. Nusbaum L, Natour J, Ferraz MB, et al. Translation, adaptation and validation of the Roland Morris questionnaire-Brazil Roland Morris. *Braz J Med Biol Res* 2001;34:203–10.
3. Feise JF, Menke M. Functional Rating Index: a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine* 2001;26:78–87.
4. Ceran F, Özcan A. The relationship of the Functional Rating Index with disability, pain, and quality of life in patients with low back pain. *Med Sci Monit* 2006;12:435–9.
5. Fairbank JCT, Couper J, Davies JB. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
6. Vernon H, Mior S. The neck disability index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991;14:409–18.
7. Bergner M, Bobbitt RA, Carter WB. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981;19:787–805.
8. Bejia I, Younes M, Kamel BS, et al. Validation of the Tunisian version of the Roland-Morris questionnaire. *Eur Spine J* 2005;14:171–4.
9. Tsang RCC. Measurement properties of the Hong Kong Chinese version of the Roland-Morris Disability Questionnaire. *Hong Kong Physiother J* 2004;22:40–9.
10. Chen SM, Liu MF, Wang BM, et al. Chinese translation and adaptation of the Roland-Morris Low Back Pain Disability Questionnaire. *FJPT* 2003;28:324–32.
11. Albert HB, Jensen AM, Dahl D, et al. Criteria validation of the Roland Morris questionnaire: a Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica. *Ugeskr Laeger* 2003;165:1875–80.
12. Brouwer S, Kuijer W, Dijkstra P, et al. Reliability and stability of the Roland Morris Disability Questionnaire. *Disabil Rehabil* 2004;26:162–5.
13. Coste J, Le Parc JM, Berge E, et al. French validation of a disability rating scale for the evaluation of low back pain (EIFEL questionnaire). *Rev Rhumatisme* 1993;60:335–41.
14. Wiesinger GF, Nuhr M, Quittan M, et al. Cross cultural adaptation of the Roland Morris questionnaire for German-speaking patients with low back pain. *Spine* 1999;24:1099–103.
15. Boscainos PJ, Sapkas G, Stilianessi E, et al. Greek versions of the Oswestry and Roland-Morris Disability Questionnaires. *Clin Orthop* 2003;411:40–53.
16. Padua R, Padua L, Ceccarelli E, et al. Italian version of the Roland Disability Questionnaire, specific for low back pain: cross cultural adaptation and validation. *Eur Spine J* 2002;11:126–9.
17. Fujiwara A, Kobayashi N, Saiki K, et al. Association of the Japanese Orthopaedic Association score with the Oswestry Disability Index, Roland Morris Questionnaire and Short Form 36. *Spine* 2003;28:1601–7.
18. Grotle M, Brox JL, Vollestad NK. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. *J Rehabil Med* 2003;35:214–47.

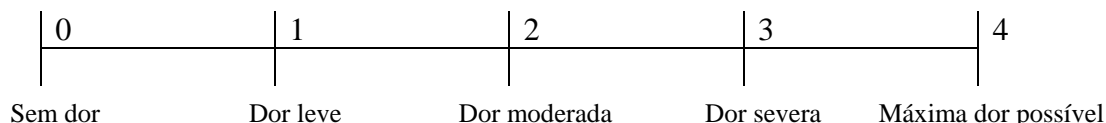
19. Kovacs FM, Llobera J, Gil Del Real MT, et al. Validation of the Spanish version of the Roland Morris questionnaire. *Spine* 2002;27:538–42.
20. Johansson E, Lindberg P. Subacute and chronic low back pain: reliability and validity of a Swedish version of the Roland and Morris Disability Questionnaire. *Scand J Rehabil Med* 1998;30:139–43.
21. Jirattanaphochai J, Jung S, Sumananont C, et al. Reliability of the Roland-Morris disability questionnaire (Thai version) for the evaluation of low back pain patients. *J Med Assoc Thai* 2005;88:407–11.
22. Kucukdeveci AA, Tennant A, Elhan AH, et al. Validation of the Turkish version of the Roland Morris Disability Questionnaire for use in low back pain. *Spine* 2001;26:2738–3743.
23. Mousavi SJ, Parnianpour M, Mehdian H, et al. The Oswestry Disability Index, the Roland Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: translation and validation studies of the Iranian versions. *Spine* 2006;31:E454–9.
24. Chansirinukor W, Maher CG, Latimer J, et al. Comparison of the Functional Rating Scale and the 18-Item Roland Morris Disability Questionnaire: responsiveness and reliability. *Spine* 2004;30:141–5.
25. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
26. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria are proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
27. Roland M, Morris R. A study of natural history of back pain: Part 1. Development of a reliable and sensitive measure of disability in low back pain. *Spine* 1983;8:141–4.
28. Deyo RA. Comparative validity of the Sickness Impact Profile and shorter scales for functional assessment in low back pain. *Spine* 1986;9:951–4.
29. Kopec KA, Esdaile JM, Abrahamowicz M. The Quebec Back Pain Disability Scale: measurement properties. *Spine* 1995;20:341–52.
30. Stratford PW, Finch F, Solomon P. Using Roland Morris scale to make decisions about individual patients. *Physiother Can* 1996;48:107–10.
31. Stratford PW, Binkley JM. A comparison study of the Back Pain Functional Scale and Roland Morris Questionnaire: North American Orthopaedic Rehabilitation Research Network. *J Rheumatol* 2000;27:1928–36.
32. Patrick DL, Deyo RA, Atlas SJ. Assessing health-related quality of life in patients with sciatica. *Spine* 1995;20:1899–908.
33. Underwood MR, Barnett AG, Vickers MR. Evaluation of two time specific low back pain outcome measures. *Spine* 1999;24:1104–12.
34. Jacob T, Baras M, Zeev A, et al. Low back pain: reliability of a set of pain measurement tools. *Arch Phys Med Rehabil* 2001;82:735–42.
35. Jensen AM, Strom SE, Turner JA, et al. Validity of the Sickness Impact Roland scale as a measure of dysfunction in chronic pain patients. *Pain* 1992;50:157–62.
36. Stratford PW, Binkley JM, Riddle DL. Development and initial validation of the Back Pain Functional Scale. *Spine* 2000;25:2095–102.
37. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 2002;82:8–24.
38. Bayar B, Bayar K, Yakut E, et al. Reliability and validity of the Functional Rating Index in older people with low back pain: preliminary report. *Aging Clin Exp Res* 2004;16:49–52.
39. Childs JD, Piva SR. Psychometric properties of the functional rating index in patients with low back pain. *Eur Spine J* 2005;14:1008–12.

Appendix 1: Brazilian – Portuguese version of FRI

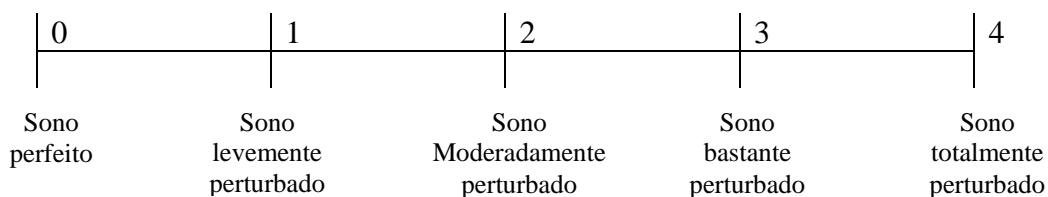
Índice de Classificação Funcional

Para avaliarmos sua condição de forma adequada, precisamos entender em até que ponto seus problemas de coluna têm afetado sua habilidade em lidar com suas atividades cotidianas. Para cada item abaixo, assinale o número que descreve sua situação neste momento de forma mais adequada.

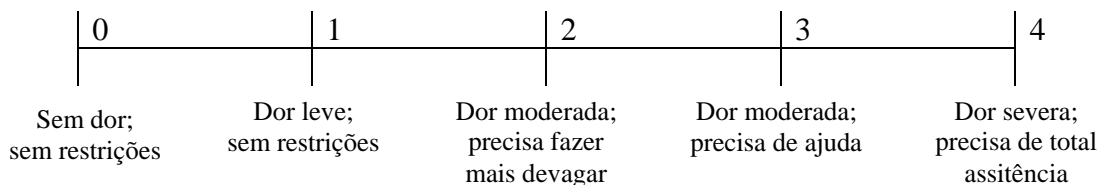
1. Intensidade da dor



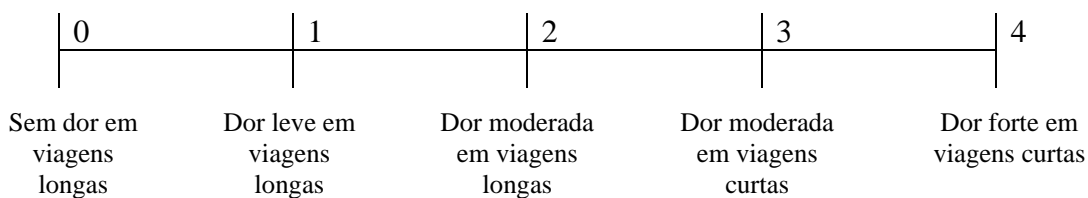
2. Sono



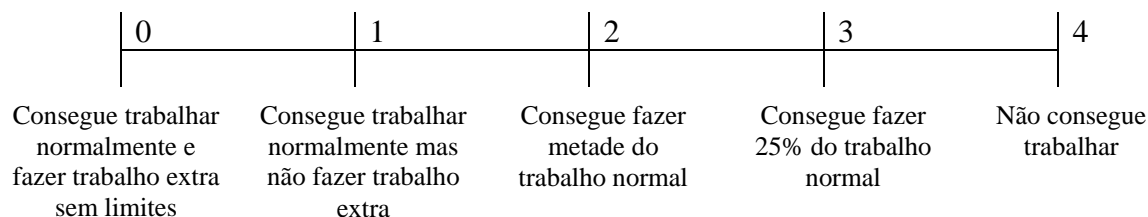
3. Cuidados pessoais (lavar-se, vestir-se, etc.)



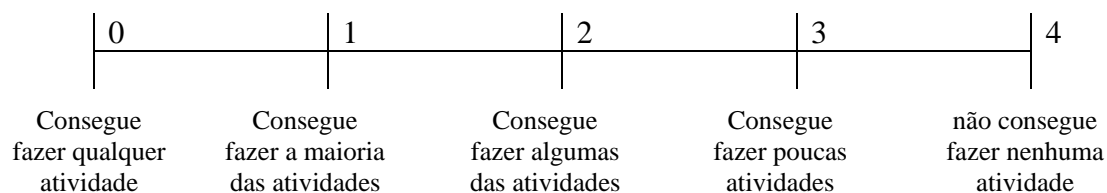
4. Viagem (dirigir, etc.)



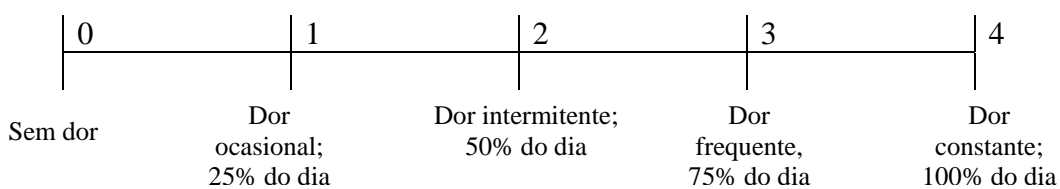
5. Trabalho



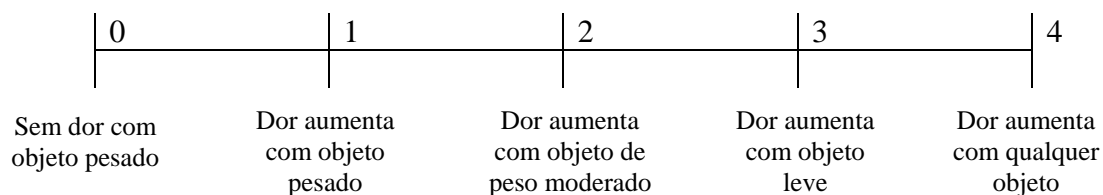
6. Lazer/recreação



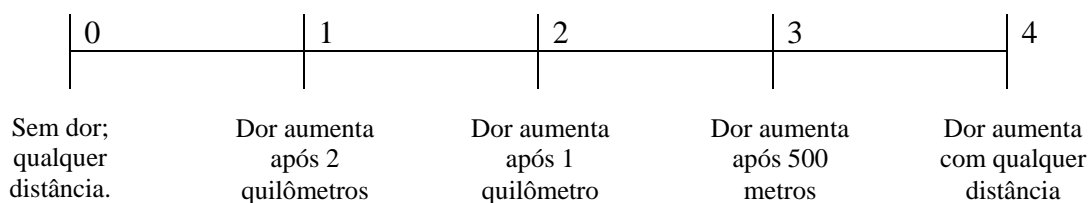
7. Frequência da dor



8. Levantamento de objetos



9. Caminhada



10. Posição de pé



Chapter Nine

Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil. Which one is the best?

Chapter Nine is published as:

Costa LOP, Maher CG, Latimer J, Ferreira PH, Ferreira ML, Pozzi GC, Freitas LMA (2008) Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil. Which one is the best? *Spine* 33:2459-2463.

**Statement from co-authors confirming authorship contribution of the
PhD candidate**

As co-authors of the paper “Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil. Which one is the best?”, we confirm that Leonardo Costa has made the following contributions:

- Conception and design of the research
- Data collection
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

Christopher G Maher		Date	<u>20.03.2009</u>
Jane Latimer		Date	<u>20.03.2009</u>
Paulo H Ferreira		Date	<u>20.03.2009</u>
Manuela L Ferreira		Date	<u>20.03.2009</u>
Giovanni C Pozzi		Date	<u>20.03.2009</u>
Ludmilla MA Freitas		Date	<u>20.03.2009</u>

Clinimetric Testing of Three Self-report Outcome Measures for Low Back Pain Patients in Brazil

Which One Is the Best?

Leonardo Oliveira Pena Costa, PT, MSc,*† Chris G. Maher, PT, PhD,* Jane Latimer, PT, PhD,*
Paulo Henrique Ferreira, PT, PhD,‡§ Manuela Loureiro Ferreira, PT, PhD,††
Giovanni Campos Pozzi, PT,† and Ludmilla Motta Andrade Freitas, PT†

Study Design. Translation, cross-cultural adaptation, and clinimetric testing of self-report outcome measures.

Objective. The aims of this investigation were to perform the translation and cross-cultural adaptation of the Patient-Specific Functional Scale (PSFS) into Brazilian-Portuguese and to perform a head-to-head comparison of the clinimetric properties of the Brazilian-Portuguese versions of the PSFS, the Roland-Morris Disability Questionnaire (RMDQ) and the Functional Rating Index (FRI).

Summary of Background Data. To date, there is no Brazilian-Portuguese version of the PSFS available and no head-to-head comparison of the Brazilian-Portuguese versions of the PSFS, RMDQ, and FRI has been undertaken.

Methods. The PSFS was translated and adapted into Brazilian-Portuguese. The PSFS, the RMDQ, and the FRI were administered to 99 patients with low back pain to evaluate internal consistency, reproducibility, ceiling and floor effects, construct validity, internal and external responsiveness. To fully test the construct validity and external responsiveness of these measures, it was necessary to cross-culturally adapt the Pain Numerical Rating Scale and the Global Perceived Effect Scale.

Results. All measures demonstrated high levels of internal consistency (Cronbach's α range = 0.88–0.90) and reproducibility (Intraclass Correlation Coefficient_{2,1} range = 0.85–0.94). High correlations among the disability-related measures were observed (Pearson's r ranging from 0.51 to 0.71). No ceiling or floor effects were detected. The PSFS was consistently more responsive than the other measures in both the internal responsiveness and external responsiveness analyses.

Conclusion. The results from this study demonstrate that the Brazilian-Portuguese versions of the RMDQ, the FRI and the PSFS have similar clinimetric properties

to each other and to the original English versions. Of all the measures tested in this study the PSFS seems the most responsive. These measures will enable international comparisons to be performed, and encourage researchers to include Portuguese speakers in their clinical trials.

Key words: low back pain, Patient-Specific Functional Scale, Roland-Morris disability questionnaire, functional rating index, Portuguese version, reproducibility, validity. **Spine 2008;33:2459–2463**

Self-report measures are widely used by clinicians and researchers for measuring the health status or outcome of treatment in patients with low back pain. Choosing the best measure is difficult given the large number of measures available. Two recently published reviews have located 36¹ and 40² self-report outcome measures for use in patients with low back pain. One sensible way of selecting a measure from this large group is to choose the one with superior clinimetric properties. Important clinimetric properties of a questionnaire are that it must be valid (*i.e.*, it measures the specific construct that it was developed to measure such as disability, pain, quality of life, *etc.*); reproducible (*i.e.*, it achieves the same results in patients whose condition remains unchanged) and responsive (*i.e.*, it detects clinically important changes over time, even if these changes are small).³

A feature of clinical measures that is frequently ignored is that the measure has to be practical for use in a busy clinical setting. Limitations of condition-specific measures of disability such as the Roland-Morris Disability Questionnaire (RMDQ)⁴ and the Oswestry Disability Index⁵ are that they contain multiple items, so are longer to complete and score, and that they can only be used for patients with low back pain. In contrast the Patient-Specific Functional Scale (PSFS)⁶ contains only 3 items, is simple to score and can be used for a wide range of health problems. The original English version of the PSFS has been shown to be valid and responsive to change in different musculoskeletal conditions, such as knee pain,⁷ cervical radiculopathy,⁸ low back pain,⁹ and neck pain.¹⁰ Additionally the PSFS has been extensively chosen as a primary outcome measure in recent randomized controlled trials.^{11–16}

Although the RMDQ,^{17,18} the Oswestry Disability Index¹⁹ and the Functional Rating Index (FRI)¹⁸ have all been cross-culturally adapted to Brazilian-Portuguese, this

From the *Musculoskeletal Division, The George Institute for International Health, School of Medicine, The University of Sydney, NSW, Australia; †Pontifícia Universidade Católica de Minas Gerais, Brazil; ‡School of Physiotherapy, The University of Sydney, Australia; and §Universidade Federal de Minas Gerais, Departamento de Fisioterapia, Belo Horizonte, Brazil.

Acknowledgment date: January 2, 2008. Revision date: May 22, 2008. Acceptance date: May 26, 2008.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Leonardo Costa is a PhD student supported by CAPES—Ministério da Educação—Brazil and Pontifícia Universidade Católica de Minas Gerais—Brazil; Chris Maher holds a research fellowship funded by the National Health and Medical Research Council of Australia.

Address correspondence and reprint requests to Leonardo Oliveira Pena Costa, PT, MSc, PO Box M201, Missenden Road, Sydney, NSW 2050, Australia; E-mail: lcos3060@gmail.com

has not yet been performed for the PSFS. Moreover, the responsiveness of Brazilian-Portuguese versions of the RMDQ and FRI were not fully tested, the original study (performed by our research group)¹⁸ measured only the internal responsiveness by calculating effect sizes (ES) and confidence intervals. Ideally, responsiveness should be tested in a follow-up study using prevalidated instruments and a global assessment scheme as an external criterion of change.²⁰

Therefore, the objectives of this investigation were to translate and cross-culturally adapt the PSFS into Brazilian-Portuguese and to perform a head-to-head comparison of the clinimetric properties of the Brazilian-Portuguese versions of the PSFS, the RMDQ, and the FRI in Brazilian patients with acute low back pain.

■ Methods

Overview of the Study Design

The study was carried out in 2 stages:

- i. The first stage was to develop a Brazilian-Portuguese translation and cross-cultural adaptation of the PSFS, the Global Perceived Effect (GPE), and the Pain Numerical Rating Scale (pain NRS);
- ii. The second stage was to test the clinimetric properties of the Brazilian-Portuguese versions of the newly developed PSFS, and the previously developed FRI, and RMDQ. This evaluation required the use of the Brazilian-Portuguese GPE (for the external responsiveness analysis) and pain NRS (for the construct validity analysis) developed in stage 1 of the study. All measures used in this study are described in Table 1 and all versions are presented in the appendix (available online through Article Plus).

Translation and Cross-Cultural Adaptation

The procedure was performed by following the recommendations from the Guidelines for the process of cross-cultural adaptation of self-report measures²¹ as follows:

1. Translation: Two independent translators translated the GPE, the Pain NRS, and the PSFS from English to Brazilian-Portuguese.
2. Synthesis of the translations: The 2 translators then synthesized the results of the independent translations and prepared the Brazilian-Portuguese versions of these measures.
3. Back translation: Using the final synthesized version, 2 new, independent translators who were totally blind to the original version translated the text back to English.
4. An expert committee comprised of the translators, health professionals and experts in the field of back pain reviewed all translations, discussed possible discrepancies and developed the final versions of the measures to be tested in Brazil.

Testing the Clinimetric Properties

This step was performed on 100 patients with new episodes of acute low back pain recruited from different physiotherapy clinics (public and private) in Belo Horizonte, Brazil. The following inclusion criteria applied: presence of low back pain for more than 24 hours but less than 6 weeks, no symptoms or treatment at least 30 days before the onset of pain, able to read, write and speak Portuguese fluently, and aged between 18 and 80 years.

The sample size was based on the advice from the report "Quality criteria for measurement properties of health status questionnaires"²² which suggests that at least 50 patients should be necessary to analyze most of the relevant clinimetric properties: internal consistency (by Cronbach's α), reproducibility, construct validity, ceiling and floor effects, and responsiveness. The quality criteria were recently developed to evaluate the design, methods, and outcomes of studies on the development and evaluation of health status questionnaires and have been used as a reference for systematic reviews of self-report outcome measures^{2,23,24} and also for cross-cultural adaptation studies.^{25,26}

The patients completed the Brazilian-Portuguese versions of the FRI, the RMDQ, the GPE, the PSFS, and the pain NRS at baseline (during the first consultation), 24-hours later (for re-

Table 1. Description of the Measures

Measure	Construct	Description
FRI ⁹	Disability	The FRI is a 10-item questionnaire with a 5-point Likert scale. The patient ranks his or her perceived disability at the present time by selecting one of the five points of the scale (0 = "no pain or full ability to function"; 4 = "worst possible pain and/or unable to perform this function at all"). The total score is obtained by summing the item scores and dividing by the maximum possible total score (40), and multiplying by one hundred percent. The range of scores is zero percent (no disability) to 100% (severe disability)
RMDQ ³	Disability	The RMDQ is a 24-item questionnaire related to normal activities of daily living. Patients are asked to circle the items, which they perceive as difficult to perform due to low back pain. Each answer is scaled either 0 (no difficulty) or 1 (difficulty), thus leaving a range of scores of 0 to 24, a higher score indicating higher disability
PSFS ¹⁰	Functional ability	In the PSFS patients are asked to identify up to three important activities that they are having difficulties with or are unable to perform due to their condition (e.g., low back pain). In addition the patients are asked to rate, on an 11-point scale (ranging from 0 "unable to perform activity" to 10 "able to perform activity at preinjury level") their current level of ability associated with each activity. The score ranges from 0 to 30, a higher score indicating higher functional ability
Pain NRS ¹¹	Pain intensity	The Pain NRS involves asking patients to rate their pain intensity levels on an 11-point scale (ranging from 0 "no pain" to 10 "pain as bad as could be"). The no. that the patient states represents his or her pain intensity score
GPE ¹²	Overall measure of change	Global perceived effect is an 11-point scale that ranges from -5 ("vastly worse") through 0 ("no change") to +5 ("completely recovered"). For all measures of global perceived effect (at baseline and all follow-ups), participants were asked "Compared to when this episode first started, how would you describe your back these days?" A higher score indicates higher recovery from the condition

producibility analysis) and after 2 weeks of treatment or discharge, whichever happened first (for responsiveness analysis). The length of the intervals (24 hours and 2 weeks for reproducibility and responsiveness, respectively) was chosen on the basis of a review of the prognosis of acute low back pain which demonstrated that patients with acute low back pain usually improve rapidly²⁷ being likely to observe improvements in 2 weeks but not in 24 hours.

Analysis

Further tests were conducted on the clinimetric properties of the adapted questionnaires. A list of the tests follows:

1. Internal consistency (homogeneity) was evaluated using Cronbach’s alpha and alpha “if item deleted”; Cronbach’s α is considered adequate between 0.70 and 0.95.²²
2. Reproducibility was tested in a test-retest design and was evaluated using the Intraclass Correlation Coefficient (ICC) (2,1). The ICC has been interpreted as follows: less than 0.40 poor reproducibility; 0.40 to 0.75 moderate reproducibility; 0.75 to 0.90 substantial and greater than 0.90 excellent reproducibility.²²
3. Construct validity was evaluated by correlating the Brazilian-Portuguese versions of the FRI, the RMDQ, an 11-point Pain NRS, the PSFS, and the GPE at baseline using Pearson’s *r*; a score of 0.70 being recommended for instruments that measure the same construct. When similar constructs are compared, scores lower than 0.70 are acceptable.²²
4. Potential ceiling and floor effects were measured by calculating the percentage of patients indicating the minimum or maximum possible scores in both questionnaires. Ceiling and floor effects are considered to be present if more than 15% of respondents achieved the lowest or highest possible total score (ceiling and floor effects are not related to individual items).²²
5. Internal responsiveness was assessed by calculating the ES with 84% CI. We calculated 84% confidence intervals for direct comparison of the ES. We chose 84% confidence intervals because nonoverlapping 84% confidence intervals are equivalent to a Z test of means at the 0.05 level.²⁸ Responsiveness indicates the sensitivity of a questionnaire to measure true change. Higher scores are preferred.²²
6. External responsiveness of the Brazilian-Portuguese versions of RMDQ, FRI, and PSFS was assessed by (i) cor-

Table 2. Characteristics of Study Participants at Baseline

Variable	
Gender	
Female (%)	57 (57.6)
Age (yrs)	44.3 (16.9)
Low back pain duration (ks)	3.45 (2.01)
Weight (kg)	72.0 (14.1)
Height (m)	1.68 (0.11)
Pain NRS (0–10)	5.69 (2.10)
Pain interference scale (1–4)	2.9 (1.15)
GPE (–5 to +5)	1.13 (1.93)
FRI (0–40)	17.28 (7.67)
RMDQ (0–24)	8.70 (5.62)
PSFS (0–30)	13.43 (5.85)

Continuous data are mean (SD), categorical data are N (%).

Table 3. Internal Consistency and Reproducibility of the Brazilian-Portuguese Versions of the FRI, RMDQ, PSFS, GPE, and Pain NRS

	Internal Consistency Cronbach’s α (Range of “ α if Item Deleted”)	Reproducibility (ICC 2,1 + 95% CI)
FRI	0.88 (0.85–0.87)	0.86 (0.77–0.95)
RMDQ	0.90 (0.88–0.90)	0.94 (0.91–0.96)
PSFS	n/a	0.85 (0.77–0.90)
GPE	n/a	0.90 (0.84–0.93)
Pain NRS	n/a	0.94 (0.90–0.96)

relating GPE rates to change scores of the Brazilian-Portuguese FRI, the Brazilian-Portuguese RMDQ and the Brazilian-Portuguese PSFS and (ii) constructing receiver operating curves using dichotomized GPE scores to categorize subjects as “improved” and “not improved.” We defined 3 different GPE cut-offs to categorize improvement, those cut-offs were: patients who scored equal or greater than 2, 3, or 4 points. The analysis is based on the area under the curve (AUC) and values around 0.70 are considered responsive.²² We compared the responsiveness of the paired AUC values using DeLong’s method.²⁹

The entire data set was double checked by 2 authors (GCP and LMAF) and a random sample (10%) was rechecked by a third author (LOPC) before the analyses.

■ Results

One hundred eligible patients were recruited and of these only 1 patient did not answer the questionnaires at all time points, leaving 99 patients in the analysis. Table 2 shows the characteristics of the study participants.

An excellent internal consistency was observed for both FRI and RMDQ with both questionnaires achieving high Cronbach’s α values (Table 3). By inspecting “alpha if item deleted” it was demonstrated that all items from the 2 questionnaires are relevant and contributed to the same construct. Additionally, the reproducibility values ranged from substantial (for the FRI and PSFS) to excellent (for the RMDQ, GPE, and Pain NRS) (Table 3). No ceiling or floor effects were detected in any of the outcome measures.

High correlations among the disability-related measures were observed, ranging from 0.51 to 0.71. Stronger correlations between disability-related measurements with the pain NRS (e.g., FRI X pain NRS $r = 0.63 P < 0.01$) than with the GPE (e.g., PSFS X GPE $r = 0.33 P < 0.01$) were found (Table 4).

Table 4. Pearson’s Correlation Amongst the Measures at Baseline

	FRI	RMDQ	GPE	PSFS	Pain NRS
FRI	1	0.71*	–0.37*	–0.53*	0.63*
RMDQ	0.71*	1	–0.42*	–0.51*	0.55*
PSFS	–0.53*	–0.51*	0.33*	1	–0.45*

* $P < 0.01$.

Table 5. Internal and External Responsiveness of the Measures

	Effect Size (84% CI)	Correlations of the Change Scores With GPE at Discharge	AUC (GPE Cut-off for Improvement = 2 or Better)	AUC (GPE Cut-off for Improvement = 3 or Better)	AUC (GPE Cut-off for Improvement = 4 or Better)
FRI	0.78 (0.68–0.87)	0.11 ($P = 0.30$)	0.46	0.57	0.57
RMDQ	0.70 (0.60–0.80)	0.14 ($P = 0.18$)	0.42	0.56	0.62
PSFS	0.95 (0.84–1.06)	0.34 ($P < 0.01$)	0.72	0.66	0.67
GPE	0.99 (0.89–1.09)	n/a	n/a	n/a	n/a
Pain NRS	1.16 (1.03–1.28)	0.02 ($P = 0.87$)	n/a	n/a	n/a

The results of internal and external responsiveness analyses are presented in Table 5. The PSFS was consistently the best measure in detecting changes over time (internal responsiveness). The PSFS ES was large, and changes in the measure were correlated with changes in the GPE scale ($r = 0.34$, external responsiveness). The PSFS also demonstrated larger AUCs, regardless of the GPE cut-off chosen, compared with the FRI and RMDQ although the AUCs did not reach statistical significance for cut-offs 3 and 4 (Table 6).

■ Discussion

The objectives of this study were to translate and cross-culturally adapt the PSFS from English to Brazilian-Portuguese and then to perform a head-to-head comparison of the clinimetric properties of the PSFS, the RMDQ and the FRI. Overall, the 3 self-report measures seem reproducible and valid; however, the responsiveness of the PSFS was consistently higher than the RMDQ and the FRI, demonstrating that the PSFS is currently the most sensitive measure available for Brazilian-Portuguese speakers.

All measures demonstrated good reproducibility evidenced by ICC (2,1) values higher than 0.80. The test-retest interval of 24 hours may have inflated our reproducibility scores; however, in patients with acute low back pain (who are likely to improve quickly) a longer time interval may allow true change in the patient's condition and lead to an error where we underestimate reproducibility scores. It was considered unlikely that a patient's condition would change substantially within 24 hours and therefore this period was chosen. Similarly, high reproducibility values have also been found for the original English³⁰ versions of the measures suggesting that these measures perform similarly across different languages and cultures.

The construct validity of the measures was confirmed by significant correlations among the PSFS, the RMDQ, the FRI, the GPE, and the pain NRS. There is a clear

association between the FRI and the RMDQ and a better association of the FRI with the pain NRS compared with the RMDQ and the PSFS, which can be attributed to the number of questions related to pain in the FRI (20% of the FRI items are associated with pain) *versus* only 1 item (out of 24) from the RMDQ and none from the PSFS.

We assessed responsiveness of the questionnaires in 2 ways. First, we evaluated internal responsiveness and found moderate ES for the RMDQ and FRI and a large ES for the PSFS. Second, we evaluated external responsiveness by calculating the AUC from the ROC curves and correlations between questionnaire scores and the GPE scores. In both cases the external responsiveness of the PSFS was higher than the RMDQ and FRI represented by the larger AUC values and higher correlations with the GPE. Interestingly, the FRI and the RMDQ were not able to distinguish between improved or not improved patients when the cut-off chosen was equal to or better than 2. However, when higher cut-offs were used, both the FRI and the RMDQ improve their AUC. This result suggests that the FRI and the RMDQ are less useful than the PSFS in measuring small improvements in a patient's condition.

The results of this study clarify some uncertainties from a previous study conducted by our group that investigated the clinimetric properties of the Brazilian-Portuguese versions of the FRI and the RMDQ in patients with chronic low back pain.¹⁸ This study reported excellent results in regards to reproducibility and validity but found poor internal responsiveness for the FRI (ES = 0.18 [84% CI 0.11–0.24]) and RMDQ (ES = 0.10 [84% CI 0.04–0.16]). A possible reason for the higher internal responsiveness of questionnaires evaluated in the current study is that patients with acute low back pain are much more likely to demonstrate true change than subjects with chronic low back pain.

Our results are similar to those reported for the original English version of the PSFS for patients with low

Table 6. Comparison of the AUC for the FRI, RMDQ and PSFS

	AUC (GPE Cut-off for Improvement = 2 or Better)	P	AUC (GPE Cut-off for Improvement = 3 or Better)	P	AUC (GPE Cut-off for Improvement = 4 or Better)	P
FRI vs. RM	0.46 vs. 0.42	0.27	0.57 vs. 0.56	0.44	0.57 vs. 0.62	0.16
FRI vs. PSFS	0.46 vs. 0.72	0.000	0.57 vs. 0.66	0.08	0.57 vs. 0.67	0.07
RM vs. PSFS	0.42 vs. 0.72	0.000	0.56 vs. 0.66	0.09	0.62 vs. 0.67	0.27

P values are from DeLong's test of paired AUC values.¹⁶

back pain.⁹ The authors performed a head-to-head comparison of the PSFS, the RMDQ, the pain NRS and physical impairment tests and found that the PSFS was consistently more responsive than the other measures (regardless of the statistics chosen). The possible reasons for the better performance of the PSFS may be due firstly, to the fact that the patient chooses the most important activities, these being activities that are more likely to change over time and secondly, that the dichotomous “yes/no” answer format of the RMDQ is less likely to be sensitive compared to the 5-point scale (from the FRI) or the 11-point PSFS scale.

The results from this study demonstrate that the Brazilian-Portuguese versions of the RMDQ, the FRI and, the PSFS have similar clinimetric properties to each other and to the original English versions.^{7,9,20,31–33} Of all the measures tested in this study the PSFS seems the most responsive. The adapted questionnaires described in this study will facilitate research on patients with low back pain who reside in countries where Portuguese is spoken. Additionally it will enable international comparisons to be performed, and encourage researchers to include Portuguese speakers in their clinical trials.

■ Key Points

- A Brazilian-Portuguese version of the Patient-Specific Functional Scale (PSFS) was developed according to the guidelines for cross-cultural adaptation.
- The Brazilian-Portuguese versions of the RMDQ, the FRI, and the PSFS have similar clinimetric properties to each other and to the original English versions.
- Of all the measures tested in this study the PSFS seems the most responsive.
- All measures seem reproducible and valid for use in clinical practice and in research studies.

Acknowledgments

The authors appreciate the help from Luciana Machado, Luciola Menezes, Marcos Noronha and Lailah de Noronha for the translation and adaptation procedures.

References

1. Grotle M, Brox JI, Vøllestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine* 2004;30:130–40.
2. Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain—searching for international cross-cultural adaptations. *Spine* 2007;32:1028–37.
3. Maher CG, Latimer J, Costa LOP. The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments. *Rev Bras Fisioter* 2007; 11:245–52.
4. Roland M, Morris R. A study of natural history of back pain, part 1: development of a reliable and sensitive measure of disability in low back pain. *Spine* 1983;8:141–4.
5. Fairbank JCT, Pynsent PB. The Oswestry disability index. *Spine* 2000;25: 2940–53.
6. Streiner DL, Norman GR. *Health Measurement Scales, a Practical Guide to Their Development and Use*. New York: Oxford University Press; 2003.
7. Chatman AB, Hyams SP, Neel JM, et al. The Patient-Specific Functional Scale: Measurement properties in patients with knee dysfunction. *Phys Ther* 1997;77:820–9.
8. Cleland JA, Fritz JM, Whitman JM, et al. The reliability and construct validity of the neck disability index and patient specific functional scale in patients with cervical radiculopathy. *Spine* 2006;31:598–602.
9. Pengel LHM, Refshauge KM, Maher CG. Responsiveness of pain, disability and physical impairment outcomes in patients with low back pain. *Spine* 2004;29:879–83.
10. Stewart M, Maher CG, Refshauge K, et al. Responsiveness of pain and disability measures for chronic whiplash. *Spine* 2007;32:580–5.
11. Pengel LHM, Refshauge KM, Maher CG, et al. Physiotherapist-direct exercise, advice or both for subacute low back pain: a randomized trial. *Ann Intern Med* 2007;146:787–96.
12. Nourbakhsh MR, Fearon FJ. The effect of oscillating-energy manual therapy on lateral epicondylitis: a randomized, placebo-control, double blinded study. *J Hand Ther* 2008;21:4–13.
13. Hancock MJ, Maher CG, Latimer J, et al. Assessment of diclofenac or spinal manipulative therapy, or both, in addition to recommended first-line treatment for acute low back pain: a randomised controlled trial. *Lancet* 2007; 370:1638–43.
14. Maher CG, Latimer J, Hodges PW, et al. The effect of motor control exercises versus placebo in patients with chronic low back pain. *BMC Musculoskeletal Disord* 2005;6:1–8.
15. Ferreira ML, Ferreira PH, Latimer J, et al. Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: a randomized trial. *Pain* 2007;131:31–7.
16. Stewart MJ, Maher CG, Refshauge KM, et al. Randomized controlled trial of exercise for chronic whiplash-associated disorders. *Pain* 2007;128:59–68.
17. Nusbaum L, Natour J, Ferraz MB, et al. Translation, adaptation and validation of the Roland Morris questionnaire—Brazil Roland Morris. *Braz J Med Biol Res* 2001;34:203–10.
18. Costa LOP, Maher CG, Latimer J, et al. Psychometric Characteristics of the Brazilian-Portuguese Versions of the Functional Rating Index and the Roland Morris Disability Questionnaire. *Spine* 2007;32:1902–7.
19. Vigatto R, Alexandre NMC, Filho HRC. Development of a Brazilian-Portuguese version of the Oswestry Disability Index: cross-cultural adaptation, reliability, and validity. *Spine* 2007;32:481–6.
20. Feise RJ, Menke MJ. Functional Rating Index—a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine* 2001;26:78–87.
21. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
22. Terwee CB, Bot SDM, Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
23. Both H, Essink-Bot ML, Busschbach J, et al. Critical review of generic and dermatology-specific health-related quality of life instruments. *J Invest Dermatol* 2007;127:2726–39.
24. Timmerman AA, Meesters CMG, Speyer R, et al. Psychometric qualities of questionnaires for the assessment of otitis media impact. *Clin Otolaryngol* 2007;32:429–39.
25. de Souza FS, Marinho CS, Siqueira FB, et al. Psychometric testing confirms that the Brazilian-Portuguese adaptations, the original versions of the Fear Avoidance Beliefs Questionnaire, and the Tampa Scale of Kinesiophobia have similar measurement properties. *Spine* 2008 33:1028–33.
26. Eisenclas JH, Harding R, Daud ML, et al. Use of the palliative outcome scale in Argentina: a cross-cultural adaptation and validation study. *J Pain Symptom Manage* 2008;35:188–202.
27. Pengel LHM, Herbert RD, Maher CG, et al. Acute low back pain: systematic review of its prognosis. *BMJ* 2003;327:323–5.
28. Tryon WW. Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical tests. *Psychol Methods* 2001;6:371–86.
29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1998;44:837–45.
30. Deyo RA. Comparative reliability of the sickness impact profile and shorter scales for functional assessment in low back pain. *Spine* 1986;9:951–4.
31. Ostelo RWJG, de Vet HCW, Knol DL, et al. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol* 2004;57:268–76.
32. Stratford PW, Gill C, Westaway M, et al. Assessing disability and change on individual patients: a report of a patient-specific measure. *Physiother Can* 1995;47:258–63.
33. Beurskens AJHM, de Vet HCW, Kōke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996; 65:71–6.

Chapter Ten

Conclusions

10.1 Principal findings

The broad aim of this thesis was to contribute to a better understanding of the contemporary management of low back pain through a series of high-quality studies. This thesis provides novel data that will influence the management of low back pain especially with regards to the efficacy of motor control exercise for patients with chronic low back pain (Chapters 2 and 3). Because a double-blind placebo-controlled design was used for the first time in a clinical trial of exercise for chronic low back pain, this study provides the least biased estimate of the effect of exercise therapy in this population. This study confirmed that while motor control exercise was an efficacious treatment option for patients with chronic low back pain, the effects of this intervention were quite small and perhaps not clinically worthwhile.

The systematic review of the reproducibility of Rehabilitative Ultrasound Imaging (RUSI) measures of abdominal wall muscles (Chapter 4) identified numerous methodological limitations in the retrieved studies, such as small sample sizes, poor patient selection and inadequate statistical analysis. Accordingly high-quality studies on the reproducibility of RUSI measures of muscle activity are urgently needed. The reproducibility study on RUSI measures for abdominal wall muscles in patients seeking care for low back pain (Chapter 5) found acceptable reproducibility for static measures, as well as for muscle activity, but not for improvement/deterioration of muscle activity in patients seeking care for low back pain.

This thesis also provides an important contribution on cross-cultural adaptation and clinimetric testing studies dedicated to low back pain questionnaires. Chapter 6 provides an overview of the topic, including information about the guidelines for

cross-cultural adaptation and methods to establish the quality of clinimetric testing, including examples. The systematic review of cross-cultural adaptations of self-report outcome measures for low back pain presented in Chapter 7 found that while there are a large number of original questionnaires available, very few have been adapted to other languages. Moreover, it found no association between the number of people who speak a language and the number of adaptations to that language. There were also considerable differences in the clinimetric testing that occurred in each adaptation. Finally two studies aimed to cross-culturally adapt and to clinimetrically test self-report outcome measures relevant to the management of low back pain in the Brazilian-Portuguese language are presented in chapters 8 and 9. It was concluded that the questionnaires chosen for these studies have similar clinimetric properties to each other and to the original English versions enabling their use in clinical practice and in research studies.

10.2 Implications and suggestions for future research

10.2.1 Motor control exercise for chronic low back pain

The findings from the randomised controlled trial presented in Chapter 3 demonstrated that motor control exercise could be recommended for patients with chronic low back pain as this intervention improved patient's function and disability in the short and long term, reduced pain intensity levels in the long term and reduced the risk of persistent pain. Although guidelines¹⁻³ already recommend exercise as an effective intervention for this population, their conclusions were based upon trials of exercise against other conservative treatments, different forms of exercise or a waiting list. Therefore the effects obtained from exercise trials are potentially overestimated

as they have not controlled for placebo effects and additional biases associated with unblinding.

Although the effects of exercise for the management of chronic low back pain reported in trials are usually classified as small or moderate at best⁴, there is always the argument that different patients will not respond similarly to the same intervention and therefore subgroups should be considered. Although this is a common (and controversial) topic of discussion, few studies have been undertaken to identify patients who may respond best to motor control exercise^{5,6} and therefore well designed studies about classification, subgrouping and clinical prediction rules are needed. For example it would be interesting to measure if the effects of motor control exercise would be different in a group of patients who have predictive characteristics of better response to the motor control exercise intervention. These characteristics were previously identified by a clinical prediction rule study⁵ which found that patients younger than 40 years, with more than 91 degrees on the straight leg raise test, with presence of aberrant movements and with a positive prone instability test were considered more likely to benefit from this intervention⁵.

Another suggestion for future studies would be to test if adding other interventions to the motor control exercise intervention would increase the size of the effect. This issue was addressed in a previous randomised controlled trial of patients with subacute low back pain that aimed to investigate the effectiveness of exercise, advice or both⁷. This study found that both exercise and advice were more effective than placebo and the effects were greater when the exercise and advice interventions were combined⁷. Another recent randomised trial aimed to investigate whether the addition

of non-steroidal anti-inflammatory drugs (NSAIDs) or spinal manipulative therapy, or both, would result in faster recovery for patients with acute low back pain receiving recommended first-line care⁸ found that the combination of spinal manipulative therapy and NSAIDs were not more effective than first line care. In addition the combined effects of NSAIDs and spinal manipulative therapy were not better than the effects of each intervention separately. Since the effects of single interventions in chronic low back pain are usually small it would be worthwhile to investigate whether the addition of other interventions over the motor control exercise would increase the size of the effect in this population.

The understanding of the underlying mechanisms on how motor control exercise works are mostly based upon biomechanical/neurophysiological studies⁹⁻¹². The ideal scenario would be to test the physiological effects of motor control exercise within an existing randomised clinical trial. In this case, physiological measures could be taken longitudinally in patients seeking care for their low back pain and a subsequent analysis with regards to the association between improvement in clinical and physiological measures would be possible. This type of study could provide important information to assist in refining the dosage, frequency and intensity of the motor control exercise program, which at the moment is firmly based on the perception of the therapist or expert opinion. An existing randomised controlled trial comparing the effects of motor control exercise against graded activity exercise¹³ which is currently in the data collection stage has been measuring physiological features such as trunk proprioception, trunk stiffness, trunk muscle response and deep muscle control. This study could potentially provide new information on the underlying mechanisms of motor control exercise.

Description of complex interventions (such as exercise regimens) has been a common problem in randomised controlled trials¹⁴. Without a detailed explanation of the interventions, the generalisability of the study is potentially compromised¹⁴. In addition it becomes a challenge for clinicians and researchers to replicate interventions presented in research manuscripts in their clinical practice and research studies. Efforts have been undertaken in order to get better descriptions of interventions; for example the revised CONSORT statement¹⁵ recommends that a comprehensive description of the intervention must be provided; this includes information about the treatment provider (including qualifications and levels of training) and how the intervention was delivered including its timing and duration. Although the motor control exercise approach is becoming increasingly popular, a clear and simple description of this intervention is still not available. The description of motor control exercise in clinical trials varies considerably¹⁶⁻¹⁸ as well as in clinical commentaries¹⁹ and textbooks^{20,21}. There is no consensus within the literature on how clinicians should plan a motor control treatment for their patients and therefore a better direction with regards to which type of exercise, frequency, progression and dosage should be established. An interesting example about description and standardisation of screening, assessment, treatment and self-management is provided by the McKenzie method. This treatment is delivered by clinicians who have received specific training on treating patients according to a classification algorithm. This algorithm also provides specific information on how the treatments should be delivered, how the treatment should be progressed and how to provide advice for patients' self-management^{22,23}.

10.2.2 Reproducibility of Rehabilitative Ultrasound Imaging measures of abdominal wall muscle activity

Chapters 4 and 5 present relevant findings with regards to the reproducibility of RUSI measures of abdominal wall muscles. Although RUSI measures of muscle thickness and muscle activity appear to be reproducible, it is not possible to accept at this stage the reproducibility of RUSI as an outcome measure to detect improvement/deterioration of muscle activity. Therefore exploratory studies testing different RUSI protocols should be undertaken in order to identify the most reproducible way of measuring clinically relevant features such as muscle activity and improvement/deterioration of muscle activity. Another challenge would be to develop RUSI tests that can be easily performed in clinical practice (without frames, load cells and complicated imaging softwares).

There is a clear need to assess the validity of RUSI measures of the deep trunk muscles. The three available studies that aimed to test the validity of RUSI measures of the activity of the abdominal wall muscles²⁴⁻²⁶ have limitations and their results are not likely to be generalizable to a clinical population. The first study aimed to test the association of muscle thickness measured by RUSI and electromyography (EMG)²⁶. A small sample of three healthy, young participants was enrolled for this study. The authors concluded that RUSI was able to detect changes in contractions of the transversus abdominis and the obliquus internus up to 20% of the maximum voluntary contraction and no relationship between the obliquus externus muscle thickness and EMG activity was found. One year later another study with similar objectives²⁴ found high correlations between muscle thickness changes and EMG activity, but similarly to the previous study, the authors recruited only 13 healthy young participants. Finally

a case-control study²⁵ conducted to compare the recruitment of the abdominal muscles by RUSI and EMG between people with and without low back pain concluded that the low back pain participants had a significantly smaller increase of the transversus abdominis thickness compared with the controls. This study enrolled 10 healthy participants and 10 patients with history of low back pain; unfortunately the “low back pain” patients were tested when they were in remission of their symptoms, and similar to the previous studies on this topic, the participants were much younger and lighter than the population seen in clinical trials. Normal issues observed in clinical populations such as pain intensity levels, disability, obesity and older age were not investigated in these studies. In summary, the available studies on the validity of RUSI measures²⁴⁻²⁶ have limitations such as lack of statistical power and imperfect participant selection and as a consequence the validity of this measure is still questionable. Therefore high quality studies in this area are urgently needed.

10.2.3 Cross-cultural adaptation and clinimetric testing of low back pain self-report outcome measures

A systematic review of international cross-cultural adaptations and clinimetric testing of self-report outcome measures relevant to the management of low back pain presented in chapter 7 provides relevant information not only in terms of the lack of availability of the most common low back pain questionnaires in non-English speaking countries, but also that the clinimetric properties of the available international versions were poorly tested. More research is needed in this area, especially in countries that have a large number of non-English speakers such as China, India and Brazil. Additionally more investigation about the clinimetric properties on the existing versions are needed, as most of them do not have important

properties tested, for example from 61 international versions, only 19 were tested for responsiveness²⁷. These studies are essential as a first step to provide outcome measures developed for testing and monitoring interventions in non-English speakers.

The results of Chapters 8 and 9 provide an important set of relevant low back pain questionnaires for patients who speak Brazilian-Portuguese; these measures were cross-culturally adapted and clinimetrically tested following the recommendations of the current guidelines on the topic^{28,29}. As a consequence clinicians and researchers from Brazil have important self-report outcome measures for their low back pain patients and research participants. Additionally multicultural English-speaking countries such as the United States, the United Kingdom and Australia could recruit Brazilian-Portuguese speakers for their clinical trials. With these questionnaires available it is now possible to increase the amount of low back pain research conducted in countries such as Brazil.

10.3 References

1. Cost B13 working group. European Guidelines for the Management of Chronic Non-Specific Low back Pain. *Eur Spine J* 2006;15 S192-S300.
2. Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: A review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:492-504.
3. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: A joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478-91.
4. Hayden JA, van Tulder MV, Malmivaara A, et al. Exercise therapy for treatment of non-specific low back pain. *Cochrane Database Syst Rev* 2005:CD000335.
5. Hicks GE, Fritz JM, Dellito A, et al. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil* 2005;86:1753-62.
6. Brennan GP, Fritz JM, Hunter SJ, et al. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain - Results of a randomized clinical trial. *Spine* 2006;31:623-31.
7. Pengel LHM, Refshauge KM, Maher CG, et al. Physiotherapist-directed exercise, advice or both for subacute low back pain: a randomized trial. *Ann Intern Med* 2007;146:787-96.
8. Hancock MJ, Maher CG, Latimer J, et al. Assessment of diclofenac or spinal manipulative therapy, or both, in addition to recommended first-line treatment for acute low back pain: a randomised controlled trial. *Lancet* 2007;370:1638-43.
9. Hodges PW, Richardson CA. Inefficient muscular stabilization of the lumbar spine associated with low back pain. A motor control evaluation of transversus abdominis. *Spine* 1996;21:2640-50.
10. Hodges PW, Richardson CA. Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Exp Brain Res* 1997;114:362-70.
11. Hodges PW, Richardson CA. Contraction of the abdominal muscles associated with movement of the lower limb. *Phys Ther* 1997;77:132-42.

12. Hides JA, Stokes MJ, Saide M, et al. Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain. *Spine* 1994;19:165-72.
13. Macedo LG, Latimer J, Maher CG, et al. Motor control or graded activity exercises for chronic low back pain? A randomised controlled trial. *BMC Musculoskelet Disord* 2008;9:65.
14. Juni P, Altman DG, Egger M. Systematic reviews in health care - Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
15. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Int Med* 2001;134:663-94.
16. Ferreira ML, Ferreira PH, Latimer J, et al. Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. *Pain* 2007;131:31-7.
17. Cairns MC, Foster NE, Wright C. Randomized controlled trial of specific spinal stabilization exercises and conventional physiotherapy for recurrent low back pain. *Spine* 2006;31:E670-E81.
18. Goldby LJ, Moore AP, Doust J, et al. A randomized controlled trial investigating the efficiency of musculoskeletal physiotherapy on chronic low back pain disorder. *Spine* 2006;31:1083-93.
19. O'Sullivan PB. Lumbar segmental 'instability': clinical presentation and specific stabilizing exercise management. *Man Ther* 2000;5:2-12.
20. Richardson CA, Jull GA, Hodges PW, et al. *Therapeutic Exercise for Spinal Segmental Stabilization in Low Back Pain*. Edinburg: Churchill Livingstone, 1999.
21. Hodges PW, Ferreira PH, Ferreira ML. Lumbar spine: Treatment of instability and disorders of movement control. In Magee DJ, Zachazewski JE, Quillen WS eds. *Scientific Foundations and Principles of Practice in Musculoskeletal Rehabilitation*: Saunders, 2007.
22. McKenzie RA, May S. *The lumbar spine: mechanical diagnosis and therapy*. Waikanae, New Zealand: Spinal Publications, 2003.
23. McKenzie R. *Treat your own back*. New Zealand: Spinal Publications, 1997.

24. McMeeken JM, Beith ID, Newham DJ, et al. The relationship between EMG and change in thickness of transversus abdominis. *Clin Biomech* 2004;19:337-42.
25. Ferreira PH, Ferreira ML, Hodges PW. Changes in recruitment of the abdominal muscles in people with low back pain - Ultrasound measurement of muscle activity. *Spine* 2004;29:2560-6.
26. Hodges PW, Pengel LH, Herbert RD, et al. Measurement of muscle contraction with ultrasound imaging. *Muscle Nerve* 2003;27:682-92.
27. Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain - Searching for international cross-cultural adaptations. *Spine* 2007;32:1028-37.
28. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186-91.
29. Terwee CB, Bot SDM, Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.

Appendices

Appendix A: Guidelines for publication for the journal *PLoS Medicine*

PLoS Medicine publishes original research articles of outstanding medical importance. We will consider manuscripts of any length; we encourage the submission of both substantial full-length bodies of work and shorter manuscripts that report novel findings that might be based on a more limited range of experiments.

The writing style should be concise and accessible, avoiding jargon so that the paper is understandable for readers outside a specialty or those whose first language is not English. Editors will make suggestions for how to achieve this, as well as suggestions for cuts or additions that could be made to the article to strengthen the argument. Our aim is to make the editorial process rigorous and consistent, but not intrusive or overbearing. Authors are encouraged to use their own voice and to decide how best to present their ideas, results, and conclusions. Although we encourage submissions from around the globe, we require that manuscripts be submitted in English. Authors who do not use English as a first language may contact us for additional information. As a step towards overcoming language barriers on acceptance of the paper, we encourage authors fluent in other languages to provide copies of their full articles or abstracts in other languages. We will publish these translations as supporting information and list them, together with other supporting information files, at the end of the article text.

Article File

Your article file should be in an editable format (DOC or RTF). We can accept PDF article files for first submissions, but if your manuscript is revised we will ask you to provide it as either a DOC or RTF file. Please note: At this time we cannot accept for review or revision any documents created in Microsoft Office 2007, even if "saved down" to the 2003 version. Major changes made in Word 2007, relative to earlier versions of Word, are incompatible with the established workflow processes of many publishers (e.g. the handling of mathematical equations). PLoS is actively seeking solutions to this problem.

Organization of the Manuscript

Most articles published in PLoS Medicine will be organized into the following sections: title, authors, affiliations, abstract, introduction, methods, results, discussion, references, acknowledgments, and figure legends. Uniformity in format will help readers and users of the journal. We recognize, however, that this format is not ideal for all types of studies. If you have a manuscript that would benefit from a different format, please contact the editors to discuss this further. Although we have no firm length restrictions for the entire manuscript or individual sections, we urge authors to present and discuss their findings concisely.

Our submission system can support a large range of formats for text and graphics, but if you experience difficulties with the site or are concerned about the suitability of your files, please contact the production department, production@plos.org.

Title (75 characters)

The title should be specific to the study yet concise, and should allow sensitive and specific electronic retrieval of the article. It should be comprehensible to readers outside your field. Avoid specialist abbreviations if possible. If the paper is a randomized controlled trial or a meta-analysis, this description should be in the title. Please also provide a brief "running head" of approximately 40 characters.

Examples: Climate Change and Increased Spread of Malaria in Sub-Saharan Africa

A Cluster-Randomized Controlled Trial of a Nurse-Led Intervention after Stroke

Please also provide a brief "running head" of approximately 40 characters.

Authors and Affiliations

Provide the first names or initials (if used), middle names or initials (if used), surnames, and affiliations—department, university or organization, city, state/province (if applicable), and country—for all authors. One of the authors should be designated as the corresponding author. It is the corresponding author's responsibility to ensure that the author list, and the summary of the author contributions to the study are accurate and complete. If the article has been submitted on behalf of a consortium, all consortium members and affiliations should be listed after the Acknowledgments. (For authorship criteria, see Supporting Information and Materials Required at Submission).

Abstract

The abstract is divided into the following three sections with these headings: Background, Methods and Findings, and Conclusions. It should contain the all following elements, except for items in square brackets, which are only needed for

some study types. Please use the same format for abstracts submitted as presubmission inquiries.

Background

This section should describe clearly the rationale for the study being done. It should end with a statement of the specific study hypothesis and/or study objectives.

Methods and Findings

Describe the participants or what was studied (eg cell lines, patient group; be as specific as possible, including numbers studied). Describe the study design/intervention/main methods used/What was primarily being assessed eg primary outcome measure and, if appropriate, over what period.

[If appropriate, include how many participants were assessed out of those enrolled eg what was the response rate for a survey.][If critical to the understanding of the paper, describe how results were analysed, ie which specific statistical tests were used.]

For the main outcomes provide a numerical result if appropriate (it nearly always is) and a measure of its precision (e.g. 95% confidence interval). Describe any adverse events or side effects. Describe the main limitations of the study.

Conclusions

Provide a general interpretation of the results with any important recommendations for future research.

[For a clinical trial provide any trial identification numbers and names (e.g. trial registration number, protocol number or acronym).

Introduction

The introduction should discuss the purpose of the study in the broader context. As you compose the introduction, think of readers who are not experts in this field.

Include a brief review of the key literature. If there are relevant controversies or disagreements in the field, they should be mentioned so that a non-expert reader can delve into these issues further. The introduction should conclude with a brief statement of the overall aim of the experiments and a comment about whether that aim was achieved.

Methods

This section should provide enough detail for reproduction of the findings. Protocols for new methods should be included, but well-established protocols may simply be referenced. Detailed methodology or supporting information relevant to the methodology can be published on our web site. This section should also include a section with descriptions of any statistical methods employed. These should conform to the criteria outlined by the Uniform Requirements, as follows: "Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as the use of P values, which fails to convey important quantitative information. Discuss the eligibility of research participants. Give details about randomization. Describe the methods for and success of any blinding of observations. Report complications of treatment. Give numbers of observations. Report losses to observation (such as dropouts from a clinical trial). References for the design of the

study and statistical methods should be to standard works when possible (with pages stated) rather than to papers in which the designs or methods were originally reported. Specify any general-use computer programs used."

Results

The results section should include all relevant positive and negative findings. The section may be divided into subsections, each with a concise subheading. Large datasets, including raw data, should be submitted as supporting files; these are published online alongside the accepted article. The results section should be written in past tense.

As outlined in the Uniform requirements, authors that present statistical data in the Results section, should "...specify the statistical methods used to analyze them.

Restrict tables and figures to those needed to explain the argument of the paper and to assess its support. Use graphs as an alternative to tables with many entries; do not duplicate data in graphs and tables. Avoid nontechnical uses of technical terms in statistics, such as "random" (which implies a randomizing device), "normal," "significant," "correlations," and "sample." Define statistical terms, abbreviations, and most symbols."

Discussion

The discussion should be concise and tightly argued. It should start with a brief summary of the main findings. It should include paragraphs on the generalisability, clinical relevance, strengths, and, most importantly, the limitations of your study. You may wish to discuss the following points also. How do the conclusions affect

the existing knowledge in the field? How can future research build on these observations? What are the key experiments that must be done.

References

Only published or accepted manuscripts should be included in the reference list.

Meetings, abstracts, conference talks, or papers that have been submitted but not yet accepted should not be cited. Limited citation of unpublished work should be included in the body of the text only. All personal communications should be supported by a letter from the relevant authors.

PLoS uses the numbered citation (citation-sequence) method. References are listed and numbered in the order that they appear in the text. In the text, citations should be indicated by the reference number in brackets. Multiple citations within a single set of brackets should be separated by commas. Where there are three or more sequential citations, they should be given as a range. Example: "...has been shown previously [1,4–6,22]." Make sure the parts of the manuscript are in the correct order for the relevant journal before ordering the citations. Figure captions and tables should be at the end of the manuscript. Because references will be linked electronically as much as possible to the papers they cite, proper formatting of the references is crucial. Please use the following style for the reference list:

Published Papers

1. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463–5467.

Please list the first five authors and then add "et al." if there are additional authors. Use of a DOI number to the full-text article is acceptable as an alternative to or in addition to traditional volume and page numbers.

Accepted Papers

Same as above, but "In press" appears instead of the page numbers. Example: Adv Clin Path. In press.

Electronic Journal Articles

1. Loker WM (1996) "Campesinos" and the crisis of modernization in Latin America. Jour Pol Ecol 3. Available: http://www.library.arizona.edu/ej/jpe/volume_3/ascii-lokeriso.txt. Accessed 11 August 2006.

Books

1. Bates B (1992) Bargaining for life: A social history of tuberculosis. Philadelphia: University of Pennsylvania Press. 435 p.

Book Chapters

1. Hansen B (1991) New York City epidemics and history for the public. In: Harden VA, Risse GB, editors. AIDS and the historian. Bethesda: National Institutes of Health. pp. 21–28.

Acknowledgments

People who contributed to the work, but do not fit the criteria for authors should be listed in the Acknowledgments, along with their contributions. You must also ensure that anyone named in the acknowledgments agrees to being so named. Details of the

funding sources that have supported the work should be confined to the funding statement. Do not include them in the Acknowledgments.

Financial Disclosure

This section should describe sources of funding that have supported the work. Please include relevant grant numbers and the URL of any funder's website. Please also include this sentence: "The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." If this statement is not correct, you must describe the role of any sponsors or funders, and amend the aforementioned sentence as needed.

Author Contributions

All authors on a paper will be contacted separately to provide information about their contribution and competing interests into our system. This information will be pulled into the article file on acceptance.

Competing Interests

This section should list specific competing interests associated with any of the authors. If authors declare that no competing interests exist, we will print a statement to this effect.

Abbreviations

Please keep abbreviations to a minimum. List all non-standard abbreviations in alphabetical order, along with their expanded form. Define them as well upon first

use in the text. Non-standard abbreviations should not be used unless they appear at least three times in the text.

Nomenclature

The use of standardized nomenclature in all fields of science and medicine is an essential step toward the integration and linking of scientific information reported in published literature. We will enforce the use of correct and established nomenclature wherever possible:

- * We strongly encourage the use of SI units. If you do not use these exclusively, please provide the SI value in parentheses after each value.

- * Species names should be italicized (e.g., *Homo sapiens*) and the full genus and species must be written out in full, both in the title of the manuscript and at the first mention of an organism in a paper; after that, the first letter of the genus name, followed by the full species name may be used.

- * Genes, mutations, genotypes, and alleles should be indicated in italics. Use the recommended name by consulting the appropriate genetic nomenclature database, e.g., HUGO for human genes. It is sometimes advisable to indicate the synonyms for the gene the first time it appears in the text. Gene prefixes such as those used for oncogenes or cellular localization should be shown in roman: v-fes, c-MYC, etc.

- * The Recommended International Non-Proprietary Name (rINN) of drugs should be provided.

Figures

If the article is accepted for publication, the author will be asked to supply high-resolution, print-ready versions of the figures. Please ensure that the files conform to our Guidelines for Figure and Table Preparation when preparing your figures for production. After acceptance, authors will also be asked to provide an attractive image to highlight their paper online. All figures will be published under a Creative Commons Attribution License, which allows them to be freely used, distributed, and built upon as long as proper attribution is given. Please do not submit any figures that have been previously copyrighted unless you have express written permission from the copyright holder to publish under the CCAL license.

Figure Legends

The aim of the figure legend should be to describe the key messages of the figure, but the figure should also be discussed in the text. An enlarged version of the figure and its full legend will often be viewed in a separate window online, and it should be possible for a reader to understand the figure without switching back and forth between this window and the relevant parts of the text. Each legend should have a concise title of no more than 15 words. The legend itself should be succinct, while still explaining all symbols and abbreviations. Avoid lengthy descriptions of methods.

Tables

All tables should have a concise title. Footnotes can be used to explain abbreviations. Citations should be indicated using the same style as outlined above. Tables occupying more than one printed page should be avoided, if possible. Larger tables

can be published as online supporting information. Tables must be cell-based; do not use picture elements, text boxes, tabs, or returns in tables. Please ensure that the files conform to our [Guidelines for Figure and Table Preparation](#) when preparing your tables for production.

Appendix B: Guidelines for publication for the journal *Physical Therapy*

Physical Therapy (**PTJ**) promotes evidence-based practice and excellence in both clinical and basic research. **PTJ** documents basic and applied knowledge related to physical therapy, provides evidence to guide clinical decision making, publishes a variety of original research relevant to the field, and is a forum for diverse opinions that are based in scholarly argument. Our readership includes physical therapist clinicians, researchers, educators, and students and all health care professionals interested in rehabilitation. Like the profession it serves, **PTJ** strives to enhance the function, health, and well-being of all members of society.

PTJ's circulation in 2008 is more than 72,000. Its 2007 impact factor was 2.152. On January 7, 2008, the mean time from submission to first decision for the previous 12 months was 47 days. Time from acceptance to publication online is less than or equal to 3 months and from acceptance to publication in print is less than or equal to 5 months. The acceptance rate is 30%.

These submission guidelines, first posted in March 2007, are MANDATORY for all manuscripts submitted on January 1, 2008, or thereafter. **PTJ** endorses the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals* put forth by the International Committee of Medical Journal Editors (ICMJE).

Editorial Policies

All submissions accepted for peer review are privileged communications. Author identity is kept confidential from reviewers, unless otherwise indicated. All correspondence is sent to the author who submits the article.

Immediate Decisions

The Editor in Chief and Editorial Board reserve the right to reject, without full review, any manuscript that does not meet Journal criteria. Manuscripts are prescreened by an Editor and/or Editorial Board member. Manuscripts are immediately rejected by **PTJ** when:

- Study participants are able bodied, without a compelling justification for relevance to physical therapist practice.
- Reliability study does not include a discussion of the impact that the error will have on clinical decision making.
- Study describes a new test without a sound comparison to current tests.
- Qualitative study is purely descriptive, without analysis or interpretation of findings.
- Study has fatal flaws in the methods section.
- Study does not have a theoretical or evidence-based argument for the relevance of the work to physical therapist practice.

Redundant, Duplicate, or Simultaneous Publication

PTJ reviews and considers a manuscript for exclusive publication with the understanding that the manuscript, or any substantial portion of the manuscript (as judged by the Editor in Chief), has not been published previously and is not under consideration for publication elsewhere, whether in print or electronic form. This policy does not usually preclude consideration of (1) a manuscript that has been rejected by another journal or (2) a complete report that follows publication of a preliminary report or pilot study. Press reports on papers presented at a scientific meeting usually will not be considered to constitute prior publication, but such reports should not be amplified by additional data or copies of tables and illustrations (see also Prior Disclosure and Media Embargo Policy).

Redundancy in publications occurs when the work of 2 or more papers from the same author overlaps substantially. Authors submitting manuscripts to **PTJ**, are asked to include in their cover letter an explanation of any prior submission or publication (eg, published article, article in press, manuscript under review, posting of results in registries, published abstract) that includes data from the same subjects studied in the submitted manuscript. Previous publication of a small fraction of the content of a manuscript does not necessarily preclude its being published in **PTJ**, but the editors need information about previous publication. Secondary analyses based on previous publications can be important scientific contributions, but the editors need to be able to judge potential redundancy. If the submitted manuscript is a report of secondary analyses, the source of the data should be properly referenced and the rationale for the secondary analysis must be provided.

In addition to the cover letter explanation, authors must provide a copy of related papers—that is, submitted or published papers that deal with the same data, in part or in full, that are reported in the manuscript being submitted to **PTJ**. These materials will be confidential and will be viewed only by the editors.

Duplicate publication is the publication of the same paper or substantially similar papers in the same journal or in more than one journal. **PTJ**'s Editorial Board holds that publication more than once of the same study results, regardless of whether the wording is the same or different, is rarely justified. Articles previously published in another language will not be considered for publication in **PTJ**.

The Editorial Board reserves the right to consult with other journals about the content of the papers in question. Furthermore, **PTJ** may decide to take one or more of the following actions:

(1) Immediately reject the manuscript without review. (2) Decline to consider any manuscripts from the author(s) for a period of time. (3) Announce publicly, within **PTJ**, that the authors have submitted a previously published article. If a paper is accepted and published before evidence of duplication is discovered, **PTJ** probably will publish a notification of redundant or duplicate publication, with or without the author's explanation or approval. Typically, **PTJ** will request that the authors write a letter acknowledging the duplicate publication. The Editorial Board may notify appropriate institutions, ranging from national databases to the authors' academic departments or university administrators, at its discretion. The Editorial Board may decide not to consider submissions from the author(s) for a period of time.

Simultaneous Submission. Authors may not submit the same manuscript simultaneously to more than one journal. If the Editorial Board learns of possible simultaneous submission, it reserves the right to consult with the other journal to which the manuscript was submitted. Furthermore, the Editorial Board may return the manuscript without review or may reject it without regard to peer reviewer recommendations and may decide not to consider any submissions from the author(s) for a period of time.

Prior Disclosure and Media Embargo Policy

New Submissions: As described in detail above, prior disclosure of any part of the contents of any manuscript in a widespread and substantive form, print or electronic, may make the manuscript ineligible for consideration by or publication in **PTJ**. Publication of abstracts and presentations at meetings do not constitute prior disclosure. Media coverage of presentations at scientific meetings will not constitute prior disclosure, as long as such coverage is not amplified by additional data or copies of tables and illustrations; however, direct release of information through press releases or news media briefings may jeopardize consideration of the submission. During the submission process, you will be prompted to indicate whether your manuscript has been presented orally at a scientific meeting or at a professional forum. Authors who need clarification of this policy are encouraged to contact the Managing Editor before releasing or distributing information from the manuscript that they want to submit.

Accepted Papers: All information regarding the content and publication date of accepted manuscripts is strictly confidential. Information contained in or about accepted articles must not appear in print, audio, video, or digital form or be released by the news media until the day before its publication date (or other specified embargo release date in the case of articles that are released early or published online ahead of print). **PTJ** is willing to work with authors who wish to present their accepted papers at conferences prior to publication. Authors are allowed to alert their university media office at the rapid proof stage but must inform **PTJ** and must ensure that the university media office observes **PTJ's** media embargo policy.

Copyright and Authorship

Authors agree to execute copyright transfer as requested during the submission process. Authors will be prompted to upload a signed copyright release, authorship, and financial disclosure form. **(Please upload one form per author.)** Manuscripts published in **PTJ** become the property of the American Physical Therapy Association (APTA) and may not be published elsewhere, in whole or in part, in print or electronic form, without the written permission of APTA, which has the right to use, reproduce, transmit, derivate, publish, and distribute the contribution, in **PTJ** or otherwise, in any form or medium.

Commercial/Financial Associations and Conflict of Interest

All funding sources supporting the work should be acknowledged. During the submission process, authors will be required to enter this information. They also will be prompted to upload a disclosure statement, signed by all authors. This information

will be held in confidence by the Editor in Chief during the review process and, if the paper is accepted for publication, will be shared with readers as appropriate.

Protection of Participants

In the cover letter that is submitted with the manuscript, authors of Research Reports should provide the name of the institutional review board (IRB), institutional animal care and use committee, or other similar body that approved the study. For those authors who do not have formal ethics review committees, the principles outlined in the Declaration of Helsinki should be followed, and authors should include a statement within the manuscript (eg, in the “Participants” section) confirming that these principles were followed. Authors also should submit patient consent forms for photographs or videos. Within the manuscript, authors must include a statement in the “Method” section that they obtained informed consent of participants, when required for protection of human subjects. Along with their signed copyright release forms, authors of Case Reports should submit a signed patient consent form. Case Report authors who practice in the United States should also include a statement about meeting the HIPAA (Health Insurance, Portability, and Accountability Act) requirements of the institution for disclosure of protected health information.

CONSORT

PTJ adopted the CONSORT (Consolidated Standards of Reporting Trials) Statement in March 2007 and the CONSORT Statement extension for nonpharmacologic treatment (NPT) interventions in February 2008. Click [here](#) for requirements.

Clinical Trials Registration

Effective January 1, 2009, **PTJ** requires clinical trial registration. Click [here](#) for details.

Photograph and Video Release

Authors must obtain and submit written permission to publish photographs or post video clips in which subjects are recognizable. This statement must be signed by the subject, parent, or guardian. You will be prompted to upload this statement during the submission process.

Reprinted Tables and Figures

Authors must obtain and submit written permission from the original sources, in the name of APTA, to publish illustrations, photographs, figures, or tables taken from those sources. Authors will be prompted to upload these permissions during the submission process.

General Instructions

The following are general instructions for preparing manuscripts for **PTJ**. Please also see specific types of manuscripts.

Related Articles

Are there other articles using the same data set or otherwise related to this manuscript that have been published or are under review by other journals? If so, please submit a masked copy of the article(s) along with your manuscript.

People-First Language

PTJ promotes "people-first" language. That is, patients and subjects should not be referred to by disability or condition (eg, use "patients who have had a stroke" or "patients with stroke," rather than "stroke patients" or "stroke survivors"). Terms that could be considered biasing or discriminatory in any way should not be used.

Scientific Writing Style

PTJ follows the *American Medical Association [AMA] Manual of Style*, 10th ed, published by Williams & Wilkins (Baltimore, Md).

Measurements

Please use the International System of Units. (English units may be given in parentheses.)

Manufacturer Footnotes

For all equipment and products mentioned in the text, place a footnote containing the manufacturer's full address with ZIP code at bottom of the page on which the item is mentioned, and use consecutive symbols (*, †, ‡, §, ||, #, **, ††, ‡‡, §§, |||, ##).

Formatting

All manuscripts must be formatted double-spaced, with pages AND lines numbered. Please use 12-point font. Most manuscripts undergo a masked review process, so you will submit both a masked copy and an unmasked copy. In the masked version, please remove author names and any affiliations within the article.

Sections, in order of appearance: (1) Title page, (2) Abstract, (3) Body of article, (4) Acknowledgments, (5) References, (6) Tables, (7) Figure legends, (8) Figures, (9) Legends for supplemental materials, (10) Appendixes.

Different article types have different requirements for word count, headings in the body of the manuscript, and number of tables and figures, please see the section on the article type for this information.

References

References should be indicated by numerical superscripts that appear consecutively in the text. If you use End Notes, please use version 6.0 or higher. References should be listed in consecutive order on a separate sheet at the end of the manuscript. Follow AMA style for reference style. Cite the reference number in the text each time an author is mentioned. Use MEDLINE/PubMed journal abbreviations. References should be listed in the order of appearance in the manuscript, not in alphabetical order.

Tables

Tables should be formatted in Word, numbered consecutively, and placed together at the end of the manuscript, after the reference list. Please refer to recent issues of **PTJ** for style.

Figures

For peer-review purposes, figures can be attached to the manuscript after the figure legends; however, figures also should be submitted as separate, high-res graphic files in tif, jpg, eps, or pdf format, with the resolution set at a minimum of 300 dpi. The separate image files will help **PTJ** staff to produce the sharpest images both in print and online. Rule of thumb: the larger the figure (eg, 8.5" × 11"), the better. If electronic formats are not available to you, figures must be submitted as 5" × 7" camera-ready glossies and mailed to the Editorial Office. Figures should be numbered consecutively. For helpful guidelines on submitting figures online, visit [Cadmus Journal Services](#). Lettering should be large, sharp, and clear, and abbreviations used within figures should agree with Journal style. Color photographs are encouraged, in sharp focus and with good contrast.

Appendixes

Appendixes should be numbered consecutively and placed at the very end of the manuscript. Use appendixes to provide essential material not suitable for figures, tables, or text. If the manuscript is accepted for publication, the review team may recommend that an appendix appear online only.

Online-Only Materials

The **PTJ** Web site has the capability to host a variety of supplemental data that cannot be published in print or that exceeds **PTJ**'s limits on word counts or on tables and figures. Supplemental files can include tables, figures, appendixes, video clips, PowerPoint files, or Excel spreadsheets.

If a manuscript contains tables or figures that exceed **PTJ**'s maximum, the review team may recommend that some of them appear online only as a PDF. These tables and figures would have the same format and style as those in the final published article.

To help the reader, **PTJ** recommends that Research Report and Case Report authors submit study protocols, treatment manuals, detailed descriptions of evaluation and intervention procedures, treatment progression algorithms, etc. These can be submitted as online-only tables, figures, appendixes, or video clips. They are reviewed by the editors and Editorial Board and should be submitted at the same time that the manuscript is submitted. The videos can be of patients, procedures, interventions, or any other relevant part of the study or case. (See [Video Central](#) for recent examples.)

Specific Format for Systematic Reviews and Meta-analyses

Guidelines and Checklists

For meta-analyses of randomized controlled trials, follow QUOROM reporting [guidelines](#) and [checklist](#). For meta-analyses of observational studies in epidemiology, follow MOOSE reporting [guidelines](#) and [checklist](#) (page 3 of PDF).

Title

Titles should not be vague and should reflect measured variables. For instance, instead of using "physical therapy" to refer to intervention, state specific interventions (eg, "strengthening exercises"). For studies that are meta-analyses or systematic reviews, add that descriptor as the subtitle at the end of the title.

Titles (including subtitles) should be no longer than **150 characters (including punctuation and spaces)**. Titles in excess of this limit will be edited, subject to author approval.

Abstracts

Word limit. 275 words

Structure. Background, Objective, Design, Setting, Patients, Intervention, Measurements, Results, Limitations, Conclusions (see [Haynes](#)).

Body of Manuscript

Word limit. 4,500 words (excluding abstract and references). Please provide the manuscript word count on the abstract page of your manuscript. Additional materials

may be submitted in the form of an appendix, which would appear only on the **PTJ** Web site if the paper is accepted.

Sections. Introduction, Methods, Results, and Discussion.

The Methods section subheadings should be:

- Data Sources and Searches
- Study Selection
- Data Extraction and Quality Assessment
- Data Synthesis and Analysis

References. 75 or fewer (See [here](#) for more information about formatting)

Tables and figures. Maximum of 8. Include a flow diagram that depicts search and selection processes; include evidence tables. (See [here](#) for more information on formatting)

Comments. Justify why the review makes an important contribution to the literature and should be a priority for publication. This is particularly critical when there are already other published reviews on the same topic or when the review locates only a few studies. Reviews that fail to provide a clear answer to the study objective are a low priority for publication.

Always end the Introduction section with a clear statement of the study's objectives or hypotheses. Outline the steps that have been taken to optimize data quality (eg, independent rating of study attributes and/or data extraction, double data entry, piloting of data extraction form, training of raters). For studies that have numerical

data and use statistical inference, include a section under Methods that describes the methods and specific statistical software used for the statistical analysis.

Where pooling is undertaken, a quantitative approach (levels of evidence, vote counting) is preferred to a qualitative approach. Include a consideration of trial heterogeneity (statistical, methodological, clinical).

Figures such as forest plots and funnel plots need to be of high quality and usually are best prepared on meta-analysis software designed to produce such plots.

Statistical analysis. Please see the recommendations in the [QUOROM](#) and [MOOSE](#) checklists for issues related to statistical analyses. Included in these recommendations are assessments of heterogeneity, study quality, and confounding.

Data. **PTJ** works to maintain the highest levels of integrity and accountability. The Editors therefore reserve the right to ask researchers to provide the raw data for their studies during review or at any time up to 5 years after publication in **PTJ**. This would likely happen only in rare instances, when credibility of the research is brought into serious question.

How to Submit a Manuscript

If you've never used [PTJ Manuscript Central](#) as an author or reviewer before, click on **Create a New Account**, and follow the prompts to submit your information and

establish a user ID and password. Once you have your user ID and password, login, click on your **Author Center**, and then click on **Submit First Draft of New Manuscript**. You will be prompted to enter data into 10 screens and then upload your manuscript.

If you're a manuscript reviewer or an author who has already used PTJ Manuscript Central, you already have a user ID and password. Login, click on **Author Center**, and either click on **Submit First Draft of New Manuscript** (if you are submitting a new manuscript) or **Revised Manuscripts** (only for those who received a manuscript decision of "Accept With Revision" or "Major Revision"). Technical assistance is available by clicking on an icon at the top of the screen; you also may contact Manuscripts Coordinator [Karen Darley](#) or Managing Editor [Jan Reynolds](#) if you have any questions.

Author Reprints

Authors are invited to order reprints of their articles. A reprint order form is sent to the corresponding author at the time of publication, along with a copy of the issue in which the article appears. Readers can contact the corresponding author of the article to obtain reprints.

Copyright © 2009 by the American Physical Therapy Association.

Appendix C: Guidelines for publication for the journal *European Spine Journal*

Types of papers

Original Articles should have no more than 2,500 words with an abstract of 150 words and 25 references. Review Articles should have no more than 3,500 words and 50-70 references. Letters to the Editor are limited to 500 words and 5 references

Manuscript Submission

Submission of a manuscript implies: that the work described has not been published before; that it is not under consideration for publication anywhere else; that its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out. The publisher will not be held legally responsible should there be any claims for compensation.

Permissions

Authors wishing to include figures, tables, or text passages that have already been published elsewhere are required to obtain permission from the copyright owner(s) for both the print and online format and to include evidence that such permission has been granted when submitting their papers. Any material received without such evidence will be assumed to originate from the authors.

Online Submission

Authors should submit their manuscripts online. Electronic submission substantially reduces the editorial processing and reviewing times and shortens overall publication

times. Please follow the hyperlink “Submit online” on the right and upload all of your manuscript files following the instructions given on the screen.

Title Page

The title page should include:

- The name(s) of the author(s)
- A concise and informative title
- The affiliation(s) and address(es) of the author(s)
- The e-mail address, telephone and fax numbers of the corresponding author

Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

Text Formatting

Manuscripts should be submitted in Word.

- Use a normal, plain font (e.g., 10-point Times Roman) for text.
- Use italics for emphasis.
- Use the automatic page numbering function to number the pages.
- Do not use field functions.
- Use tab stops or other commands for indents, not the space bar.
- Use the table function, not spreadsheets, to make tables.

- Use the equation editor or MathType for equations.

Note: If you use Word 2007, do not create the equations with the default equation editor but use the Microsoft equation editor or MathType instead.

- Save your file in doc format. Do not submit docx files.
- Manuscripts with mathematical content can also be submitted in LaTeX.

Headings

Please use no more than three levels of displayed headings.

Abbreviations

Abbreviations should be defined at first mention and used consistently thereafter.

Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols. Always use footnotes instead of endnotes.

Acknowledgments

Acknowledgments of people, grants, funds, etc. should be placed in a separate section before the reference list. The names of funding organizations should be written in full.

Citation

Reference citations in the text should be identified by numbers in square brackets.

Some examples:

1. Negotiation research spans many disciplines [3].
2. This result was later contradicted by Becker and Seligman [5].
3. This effect has been widely studied [1-3, 7].

Reference list

The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text. Do not use footnotes or endnotes as a substitute for a reference list. The entries in the list should be numbered consecutively.

- **Journal article**

Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L (2009) Effect of high intensity intermittent training on heart rate variability in prepubescent children. *Eur J Appl Physiol* 105:731-738. doi: 10.1007/s00421-008-0955-8

Ideally, the names of all authors should be provided, but the usage of “et al” in long author lists will also be accepted:

Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. *N Engl J Med* 965:325–329

- **Article by DOI**

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. *J Mol Med*. Doi:10.1007/s001090000086

- **Book**

South J, Blass B (2001) *The future of modern genomics*. Blackwell, London

- **Book chapter**

Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) *The rise of modern genomics*, 3rd edn. Wiley, New York, pp 230-257

- **Online document**

Doe J (1999) Title of subordinate document. In: *The dictionary of substances and their effects*. Royal Society of Chemistry. Available via DIALOG.
[http://www.rsc.org/dose/title of subordinate document](http://www.rsc.org/dose/title%20of%20subordinate%20document). Accessed 15 Jan 1999

- **Dissertation**

Trent JW (1975) *Experimental acute renal failure*. Dissertation, University of California

Always use the standard abbreviation of a journal's name according to the ISSN List of Title Word Abbreviations, see www.issn.org/2-22661-LTWA-online.php

Tables

- All tables are to be numbered using Arabic numerals.
- Tables should always be cited in text in consecutive numerical order.

- For each table, please supply a table caption (title) explaining the components of the table.
- Identify any previously published material by giving the original source in the form of a reference at the end of the table caption.
- Footnotes to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data) and included beneath the table body.

Artwork Guidelines

For the best quality final product, it is highly recommended that you submit all of your artwork – photographs, line drawings, etc. – in an electronic format. Your art will then be produced to the highest standards with the greatest accuracy to detail.

The published work will directly reflect the quality of the artwork provided.

- Figures should always be cited in text in consecutive numerical order.
- Figure parts should be denoted by lowercase letters (a, b, c, etc.).
- If an appendix appears in your article and it contains one or more figures, continue the consecutive numbering of the main text. Do not number the appendix figures, "A1, A2, A3, etc." Figures in online appendices (Electronic Supplementary Material) should, however, be numbered separately.

Figure Captions

- Each figure should have a concise caption describing accurately what the figure depicts. Include the captions in the text file of the manuscript, not in the figure file.

- Figure captions begin with the term **Fig.** in bold type, followed by the figure number, also in bold type.
- No punctuation is to be included after the number, nor is any punctuation to be placed at the end of the caption.
- Identify all elements found in the figure in the figure caption; and use boxes, circles, etc., as coordinate points in graphs.
- Identify previously published material by giving the original source in the form of a reference citation at the end of the figure caption.

Figure Placement and Size

- When preparing your figures, size figures to fit in the column width.
- For most journals the figures should be 39 mm, 84 mm, 129 mm, or 174 mm wide and not higher than 234 mm.
- For books and book-sized journals, the figures should be 80 mm or 122 mm wide and not higher than 198 mm.

Permissions

If you include figures that have already been published elsewhere, you must obtain permission from the copyright owner(s) for both the print and online format. Please be aware that some publishers do not grant electronic rights for free and that Springer will not be able to refund any costs that may have occurred to receive these permissions. In such cases, material from other sources should be used.

Integrity of research and reporting

Conflict of interest

Authors must indicate whether or not they have a financial relationship with the organization that sponsored the research. They should also state that they have full control of all primary data and that they agree to allow the journal to review their data if requested. Therefore the manuscript must be accompanied by the “Conflict of Interest Disclosure Form”. To download this form, please follow the hyperlink on the right.

Copyright information

Submission of a manuscript implies: that the work described has not been published before (except in the form of an abstract or as part of a published lecture, review, or thesis); that it is not under consideration for publication elsewhere; that its publication has been approved by all coauthors, if any, as well as by the responsible authorities at the institute where the work has been carried out; that, if and when the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the publisher and that the manuscript will not be published elsewhere in any language without the consent of the copyright holders.

All articles published in this journal are protected by copyright, which covers the exclusive rights to reproduce and distribute the article (e.g., as offprints), as well as all translation rights. No material published in this journal may be reproduced photographically or stored on microfilm, in electronic data bases, video disks, etc., without first obtaining written permission from the publisher.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if not specifically identified, does not imply that these names are not protected by the relevant laws and regulations.

While the advice and information in this journal is believed to be true and accurate at the date of its going to press, neither the authors, the editors, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Online First publication

Online First articles are published in electronic form weeks before distribution of the printed journal. An Online publication date is published in each Online First article and its print version. Authors should be aware that after electronic publication they cannot withdraw an article or change its content. Any corrections have to be made in an Erratum which will be hyperlinked to the article. Online First articles can be cited using the Digital Object Identifier, a unique and consistent identification code included in both the print and the electronic versions.

Special regulations for photocopies in the USA. Photocopies may be made for personal or in-house use beyond the limitations stipulated under Section 107 or 108 of U.S. Copyright Law, provided a fee is paid. All fees should be paid to the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, USA, stating the ISSN 0940-6719, the volume, and the first and last page numbers of each article copied. The copyright owner's consent does not include copying for general distribution, promotion, new works, or resale. In these cases, specific written permission must first be obtained from the publisher.

The Canada Institute for Scientific and Technical Information (CISTI) provides a comprehensive, world-wide document delivery service for all Springer-Verlag journals. For more information, or to place an order for a copyright-cleared Springer-Verlag document, please contact Client Assistant, Document Delivery, Canada Institute for Scientific and Technical Information, Ottawa K1A 0S2, Canada (Tel: 613-993-9251; Fax: 613-952-8243; e-mail: cisti.docdel@nrc.ca).