Chronic Low Back Pain: Diagnostic Classification, Physical Assessment and

Ultrasound Imaging

Thesis submitted by

Carol Ann Flavell, MSc, GDip Phty

for the degree of Doctor of Philosophy

Discipline of Physiotherapy

in the College of Healthcare Sciences

James Cook University

December 2017

Acknowledgements

Without the constant support, untiring hard work, guidance and constructive feedback of Professor Susan Gordon, this research thesis would not have been possible. I cannot thank Susan enough for the guidance and inspiration she has provided over the past six years.

Whole-hearted appreciation goes to my co-advisor Associate Professor Laurence Marshman who, despite a busy clinical caseload, provided a focused, dependable and collaborative cross-discipline perspective to this research. Likewise, my appreciation goes to Associate Professor Kerrianne Watt for her statistics advice, which helped bring this project to fruition. Thanks go to bioengineers Dr Brian Anthony and Dr Matthew Gilbertson from Massachusetts Institute of Technology, who developed the new technology used in this project and provided ongoing advice, and also to the assistants who randomised and blinded data, and the nursing and administrative staff of Queensland Health and Mater Health Services.

To my partner Paul Gabriel, who has supported me through high and low times. Thanks for sacrificing your needs so that I could keep on track. Thanks also to my sons Luke and Joshua, who always understood and respected why this was so important to me. You are both a constant joy in my life—likewise, my siblings in England, who have given their support every step of the way. Finally, my appreciation to James Cook University for awarding me a Higher Degree by Research Competitive Scholarship, which supported this project during its final stages.

Statement of the Contribution of Others

Assistance	Contributors	Affiliations
Advisory panel	Professor Susan Gordon (primary advisor)	Flinders University and James Cook University Adjunct
	Associate Professor Laurence Marshman	James Cook University Medical School; Townsville Hospital; and Mater Hospital, Pimlico
	Associate Professor Kerrianne Watt	James Cook University, Public Health
Doctoral cohort	Dr Melissa Crowe	
	Dr Jenni Judd	James Cook University
	Dr Diana Mendez	
Financial assistance	Research Infrastructure Block Grants of $36,000 and $13,332.37	
	Physiotherapy Seeding Grant Fund $8000	James Cook University
	James Cook University HDR Competitive Scholarship	
Bioengineering advice and support	Dr Brian Anthony	Massachusetts Institute of Technology
	Dr Matthew Gilbertson	
Data collection and extrapolation	Townsville Hospital	Queensland Health
	Mater Hospital, Pimlico	Mater Health Services
	Rehabilitation Sciences technical staff	James Cook University, College of Healthcare Sciences

Contribution Chapter	Publication Title	Contribution
Chapter 3	'Inter-Rater Reliability of Classification Systems in Chronic Low Back Pain Populations'	Concept, Literature Search: CAF Literature Synthesis and Independent Appraisal: CAF/SJG Manuscript Writing, Figures and Tables: CAF Draft/Proof Reading: CAF/SJG/LAGM/KW
Chapter 5	'Intra-Examiner Reliability of Lumbar Spine and Neuro-Dynamic Flexibility Measurements in an Older and Overweight Healthy Asymptomatic Population'	Concept: CAF Ethics Application: CAF/SJG Data Collection: CAF Data Analysis: CAF/KW Manuscript Writing, Figures and Tables: CAF Draft/Proof Reading: CAF/SJG/KW
Chapter 6	'Classification Characteristics of a Chronic Low Back Pain Population Using a Combined Mckenzie and Patho-Anatomical Assessment'	Concept: CAF Ethics Application: CAF/SJG/LAGM Data Collection: CAF/LAGM Data Analysis: CAF Manuscript Writing, Figures and Tables: CAF Draft/Proof Reading: CAF/SJG/LAGM
Chapter 7	'Reply to the Letter to the Editor Regarding "Classification Characteristics of a Chronic Low Back Pain Population Using a Combined Mckenzie and Patho-Anatomical Assessment"'	Response Writing: CAF Draft/Proof Reading: CAF/SJG/LAGM

CAF, Carol A. Flavell; SJG, Susan J. Gordon; LAGM, Laurence A. G. Marshman; KW, Kerrianne

Watt

12th December 2017

Carol Ann Flavell   Date

Abstract

Optimal physiotherapy examination and the resulting classifications provided for chronic low back pain (CLBP) patients should involve the application of evidence-based methods. An ideal examination and classification process will be valid and reliable, with demonstrated improvement in patient outcomes to evidence its overall worth. To this end, valid examination processes for diagnostic classification, and established reliability of physical assessment and clinical measurement, are required. This thesis aimed to develop an evidence-based physiotherapy examination algorithm to classify CLBP, to establish reliable methods for clinical measurement of lumbar spine range of motion, and to trial a novel standardised real-time ultrasound imaging method to measure the transversus abdominis in CLBP.

A physiotherapy CLBP examination algorithm (MK-C), with reliable methods of clinical measurement, were applied in a series of studies. Studies were conducted in a CLBP population to determine classification characteristics, and diagnostic agreement between a physiotherapist and available reference standards. Additionally, standardised and reliable real-time ultrasound imaging methods were applied to investigate the function of the transversus abdominis following pain abolition in CLBP.

Literature reviews identified evidence-based components for inclusion in the MK-C. Two systematic reviews appraised CLBP studies by using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses methods. The first appraised validity or diagnostic accuracy of physiotherapy low back pain classification systems, individual examination tests or test clusters ($N = 5$). Clustered clinical tests for radiculopathy, discogenic pain and facet joint syndrome had been

validated in CLBP. The second review appraised reliability of physiotherapy low back pain classification systems ($N = 3$). High risk of bias and variable inter-rater reliability ($k = .32$ to $.96$) were identified. There was no evidence to suggest that any existing physiotherapy low back pain classification system had demonstrated high reliability or reference standard validity as a standalone method of examination in a CLBP population.

Intra-examiner reliability of lumbar spine range of motion and joint range of motion associated with lumbar neuro-dynamic tests was reported in a blinded test–retest study of asymptomatic participants ($N = 19$) with demographics that simulated CLBP patients (age $\geq$ 50 years; body mass index $\geq$ 24). High intra-examiner measurement reliability ($ICC = .68$–$.99$) was established using standard tape measure and goniometry protocols. This supported the suitability of these measurements in CLBP examination.

The MK-C was applied to classify CLBP and report characteristics within and between classification categories by using a cross-sectional study design ($N = 150$). Reported characteristics were age, gender, body mass index (BMI), pain intensity scored on a visual analogue scale, pain duration (months), disability scored on the Oswestry Disability Index and Roland–Morris disability questionnaire, and pain somatisation rated using the Modified Somatic Pain Perceptions Questionnaire. Results supported utility of the MK-C for CLBP examination, with 94% of participants classified at first attendance. Facet joint syndrome was most frequently classified. All classification categories demonstrated 'distressed' levels of pain somatisation (score $\geq$ 13). Age, disability and pain somatisation were distinguishing CLBP characteristics.

A second blinded cross-sectional study ($N = 92$) reported level of CLBP diagnostic agreement between a physiotherapist using the MK-C and available reference standards, which included a CLBP medical specialist's diagnosis based on clinical examination and magnetic resonance imaging, and the outcome of a diagnostic spinal anaesthetic injection when clinically indicated. Observed examiner agreement was 51%, with **'Fair'** after-chance agreement ($k = .22$, CI [1.57, −1.13]). Chi-square analysis of subjects who received a diagnostic spinal anaesthetic injection ($N = 50$) identified that combined examiner diagnosis was truly positive in 46% of cases.

A blinded intra-examiner reliability study applied standardised probe force, inclination and roll using a 'force probe device' during transversus abdominis real-time ultrasound imaging in CLBP ($N = 17$). High measurement reliability was reported for resting transversus abdominis ($ICC = .98$, CI [0.93, 0.99]), contracted transversus abdominis ($ICC = .99$, CI [0.97, 0.99]) and transversus abdominis activation ($ICC = .93$, CI [0.82, 0.97]). This was superior to previous reliability reported using 'free-hand' real-time ultrasound imaging.

A pre–post intervention study ($N = 47$) used the same ultrasound imaging methods to measure transversus abdominis activation before and following pain relief from a diagnostic spinal anaesthetic injection (≥ 75% pre–post reduction on a visual analogue scale). Transversus abdominis activation was calculated as a rest to contracted thickness change from paired images ($N = 324$). Results indicated that pain relief did not immediately result in a statistically ($p < .05$) or clinically (≥ 20%) significant improvement in transversus abdominis activation. Thus, as an isolated intervention, pain relief appears insufficient, supporting the need for transversus abdominis retraining following CLBP relief.

Clinical contributions:

- This study has provided a unique contribution to knowledge on diagnostic classification, physical assessment and real-time ultrasound imaging in CLBP.

- New knowledge has been presented about physiotherapy CLBP diagnostic classification characteristics and agreement using an MK-C examination algorithm, and transversus abdominis function using standardised real-time ultrasound imaging methods. This will assist treatment focus and CLBP resource planning.

- Intra-rater reliability of established physical assessment measurements has been confirmed for lumbar spine range of movement in CLBP to guide clinical measurement during physiotherapy examination.

- Characteristics of CLBP and between diagnostic classification categories have been reported, which provide knowledge to guide clinical reasoning.

- MK-C 'Fair' diagnostic agreement, although not optimal, provides clinicians with new knowledge on the value of the MK-C, which can be considered an evidence-based architype examination algorithm suitable to classify CLBP.

- Pain abolition in isolation does not enhance transversus abdominis activation, suggesting that routine physiotherapist-guided transversus abdominis retraining immediately following pain abolition interventions may be indicated. Future research should determine if this approach improves patient outcomes.

Research implications:

- A critical need for research specific to physiotherapy examination and classification in CLBP has been highlighted.

- Research evidence related to physiotherapy examination of CLBP has been provided, which can be used as a comparator for future studies.

- Improved and reliable methods of real-time ultrasound imaging for transversus abdominis measurement using a 'force probe device' have been demonstrated, which can be used to improve measurement accuracy in future studies.

- Research evidence about the complex relationship between pain abolition and transversus abdominis function in CLBP has been provided, which was previously subject to hypothetical assumption.

Table of Contents

List of Tables

List of Figures

List of Plates

List of Abbreviations

+ve	Positive
−ve	Negative
ADIM	Abdominal draw-in manoeuvre
ANR	Adherent nerve root
BMI	Body mass index
CI	Confidence interval
CLBP	Chronic low back pain
CP	Centralisation phenomenon
CPR	Clinical prediction rule
CT	Computerised tomography
CTrA	Transversus abdominis thickness contracted
DP	Directional preference
DSAI	Diagnostic spinal anaesthetic injection
ER	Extension rotation test
FGFJI	Intra-articular lumbar facet joint injection
FJS	Facet joint syndrome
FPD	'Force probe device'
ICC	Intra-class correlation coefficient
IQR	Interquartile range
LBP	Low back pain
LR	Likelihood ratio
Mdn	Median
MII	McKenzie Institute International
MRI	Magnetic resonance imaging

MSI	Movement system impairment classification
MSPQ	Modified Somatic Pain Perceptions Questionnaire
NCD	Neurosurgeon clinical diagnosis
NR	Nerve root
OCS	Motor control impairment classification system
ODI	Oswestry Disability Index
PAIVM	Passive accessory intervertebral movement
PSLR	Passive straight leg raise
QAREL	Quality Appraisal of Reliability Studies
QUADAS	Quality Assessment of Comparative Diagnostic Accuracy Studies
RM	Roland–Morris disability questionnaire
ROC	Receiver operating characteristic
ROM	Range of motion
RTrA	Transversus abdominis thickness at rest
RTUI	Real-time ultrasound imaging
SD	Standard deviation
SEM	Standard error of measurement
SI	Spinal instability
SIJS	Sacroiliac joint syndrome
Sn	Sensitivity
Sp	Specificity
TFNRB	Transforaminal nerve root block
TrA	Transversus abdominis
TrA-C	Transversus abdominis activation
VAS	Visual analogue scale

Chapter 1: Introduction

1.1 Chronic Low Back Pain

1.1.1 Background

Low back pain (LBP) is defined as pain between the level of the 12th thoracic vertebra and the buttock crease, with or without associated lower limb symptoms (Jones, Watson, Silman, Symmons & Macfarlane, 2003). It is a common condition, experienced by 70–80% of the population at some period of their life (Arab, Rasouli, Amiri, & Tahan, 2013; Tahan, Rasouli, Arab, Khademi, & Samani, 2014), with a higher prevalence in women and people aged 40–80 years (Hoy et al., 2012).

Although the majority of individuals will experience LBP symptoms for less than 1 month (Pengel, Herbert, Maher, & Refshauge, 2003), some experience recurrent symptoms (Deyo et al., 2014) and Hong, Reed, Novick, and Happich, (2013) reported that approximately 10–20% of people will develop chronic low back pain (CLBP). By definition, CLBP, a subgroup of LBP, is persistent for 12 weeks or more. It is unremitting and more complex than other LBP subgroups, and results in functional loss, disability, and emotional, financial and social cost for affected individuals and society, and may also adversely affects individuals' relationship networks (Guclu, Guclu, Ozaner, Senormanci, & Konkan, 2012; Prins, van der Wurff, & Groen, 2013; Reneman et al., 2006; Tracey & Bushnell, 2009).

1.1.2 Prevalence and aetiology

CLBP has been recognised internationally as a condition of increasing prevalence and financial cost (Freburger et al., 2009; Hong et al., 2013). In 2014–15, approximately 3.7 million Australians had chronic back problems, accounting for 1.8% of total government healthcare expenditure (Australian Institute of Health and Welfare, 2016).The diverse aetiology of CLBP can be related to factors that include

patho-anatomical disorders (Ali et al., 2013; Boswell et al., 2015; Eirikstoft & Kongsted, 2014; Kuslich, Ulstrom, & Michael, 1991; Laslett & van Wijmen, 1999; Long et al., 1996; O'Sullivan, 2000; Petersen et al., 2003), psychosocial influences (Moore., 2010), lumbar dyskinesia and motor control impairment (O'Sullivan, 2005). The complexity of this condition results in any or many of these factors being a cause or consequence of CLBP (Ali et al., 2013; Boswell et al., 2015; Eirikstoft & Kongsted, 2014; Kuslich, Ulstrom, & Michael, 1991; Laslett & van Wijmen, 1999; Long et al., 1996; Moore., 2010; O'Sullivan, 2000; O'Sullivan, 2005; Petersen et al., 2003). Additionally, sinister 'red flags' of spinal, pelvic or visceral pathology may masquerade as CLBP. Anatomical structures that contribute to CLBP include the facet joints (Boswell et al., 2015), intervertebral discs and associated vertebral endplates (Ali et al., 2013) and ligamentous or myofascial structures (Kuslich, Ulstrom, & Michael, 1991), any or all of which may lead to spinal instability (Eirikstoft & Kongsted, 2014; Laslett & van Wijmen, 1999; Long et al., 1996; O'Sullivan, 2000; Petersen et al., 2003). Although not specific to the lumbar spine, the sacroiliac joint, with symptom distribution to the buttock, may also masquerade as pain of lumbar spine origin.

In combination or isolation, aetiological factors contribute to the functional loss evident in CLBP patients. Hence, effective treatment requires comprehensive initial and ongoing examination methods to identify specific contributing factors. Multidisciplinary patient assessment and treatment in specialist CLBP management clinics is advocated, consisting of a multidisciplinary team of specialised doctors, nurses, occupational therapists, physiotherapists and psychologists (Kamper et al., 2015). Regrettably, few of these clinics currently exist (Kamper et al., 2015). Consequently, the ideal multidisciplinary team approach in specialised CLBP

2

management is rare, with physiotherapy for CLBP often conducted separately in private practices and hospital physiotherapy departments.

Regardless of the limited availability of designated CLBP management clinics or the physiotherapist's clinical setting, physiotherapy examination and classification of CLBP should be based on valid and reliable processes. These should be specific to the target population to guide efficacious treatment.

1.2 Physiotherapy Clinical Examination and Classification of Chronic Low Back Pain

1.2.1 Background to clinical examination

Optimal physiotherapy examination and the resulting classifications provided for CLBP patients should involve the application of evidence-based methods. An ideal examination and classification process will be valid and reliable, with demonstrated improvement in patient outcomes to evidence its overall worth (Saragiotto, Maher, Hancock, & Koes 2017). To this end, valid examination processes for diagnostic classification, and established reliability of physical assessment and clinical measurement are required.

Evidence-based physiotherapy examination using manual clinical examination techniques to identify an individual's physical limitations and functional loss remains the underpinning foundation for physiotherapy assessment, and ultimately leads to a classification of the patients CLBP, and informs effective treatment planning.

The use of manual therapy to treat spinal pain predates the birth of modern medicine (Paris, 2000). In the early 20th century, doctors such as James Mennell and Edgar Cyriax fostered the use of manual therapy by physiotherapists for examination and treatment of LBP. Their work was continued by their sons, who further advanced

the use of manual skills within the physiotherapy profession (Paris, 2000). In the 1950s and 60s, the Australian and New Zealand physiotherapy professions witnessed pioneer contributions to manual therapy for spinal examination by Geoff Maitland, Robin McKenzie, John Mennell and Stanley Paris (International Maitland Teacher's Association, 2010; May, 2013; Paris, 2000). The subsequent rise in the popularity of manual therapy coincided with a transition from physiotherapy LBP management, based on a medical practitioner's prescription of heat, massage and exercise, to the first-contact autonomous physiotherapy practice of today. Accordingly, contemporary physiotherapy management of LBP is guided by detailed examination processes, with the application of diagnostic skills and clinical reasoning (Petty, 2013). Examination includes a subjective component, defined by questioning and history taking, and an objective component defined by physical tests and assessments (Petty, 2013).

A variety of physiotherapy examination or classification systems are currently used to identify physical impairments, movement restriction and diagnose the source of an individual's LBP (Hill et al., 2008; McKenzie, 1981; O'Sullivan, 2005; Petersen et al., 2003; Sahrmann et al., 2003). However, anatomical complexity challenges physiotherapy examination of the lumbar spine and diagnosis of associated pathologies. In CLBP, these challenges are compounded by the multifaceted interaction of biological, social and behavioural factors, as well as individual characteristics specific to a given classification category. These may present alone or in combination (Deyo et al., 2014). The examination system a physiotherapist uses depends on preference, clinical exposure and specific training (Petersen et al., 2003), but ultimately the examination process should be specific to the target population and evidence based, ensuring optimal validity and reliability.

4

Because LBP is not a homogeneous condition, results reported from studies conducted on heterogenic LBP populations, should not be extrapolated to CLBP.

1.2.2 Physiotherapy classification compared with medical diagnosis

It is important that CLBP examination conclusions drawn from physiotherapy assessment align with those of doctors and other health providers. This is because chronic pain management is optimal when examination conclusions and diagnostic classifications across multiple health disciplines are considered in the overall management decision for each CLBP patient (Stanos, 2012).

However, a medical specialist's examination differs from that of a physiotherapist, because the former can request and conduct advanced diagnostic investigations (Boswell et al., 2015). In accordance with CLBP clinical management guidelines, a medical specialist combines the patient's current clinical symptoms, past medical history and physical examination findings with advanced diagnostic investigations and procedures.

One such diagnostic investigation is magnetic resonance imaging (MRI). MRI is a relatively safe imaging modality that uses non-invasive magnetic fields to image soft tissues and intra-spinal structures (Brown, Cheng, & Haacke, 2014; Chou et al., 2007). It has excellent soft-tissue discrimination and is the most sensitive modality for spinal evaluation (A. Lamont, personal communication, 6 October 2015). However, clinical examination and MRI may sometimes prove unsuccessful to diagnose a specific source of CLBP, because frequently there is coincidental presence of unrelated spinal anomalies in patients with LBP (Jensen et al., 1994). In these cases, following a clinical examination and MRI review, a symptomatic patho-anatomical source may be determined or refuted using more invasive procedures, such as diagnostic spinal anaesthetic injection (DSAI)

5

conducted by an appropriately trained medical specialist (Curatolo & Bogduk, 2010; Ngan, Tuan, Son, Minh, & Dey, 2016; Seising, 2006).

1.2.3 Medical diagnosis confirmed by diagnostic spinal anaesthetic injection

In 1901, a spinal injection procedure that involved instillation of cocaine into or around spinal structures to reduce sciatica was reported (Nelson & Landau, 2001). By 1925, sacral epidurals were being conducted using a local anaesthetic called procaine, and by 1936, following the discovery of cortisone, steroid injections became widespread (Nelson & Landau, 2001). Since the 1950s, injection types and drug instillation protocols have diversified, and the ability of spinal injection to reduce or abolish symptoms provides diagnostic efficacy to rule in or out a specific patho-anatomical source of LBP symptoms (Ngan et al., 2016; Seising, 2006).

Spinal injection procedures performed for diagnostic purposes include intra-articular lumbar facet joint injection (FGFJI), sacroiliac joint injection, transforaminal nerve root block (TFNRB) conducted under fluoroscopy guidance and provocation discography. Such procedures are useful diagnostic tools to supplement clinical examination and diagnostic imaging, as they allow localisation of a symptomatic structure that may be inaccessible via clinical examination (Curatolo & Bogduk, 2010).

FGFJI for the diagnosis of facet joint syndrome (FJS), and TRNRB to diagnose radicular pain or nerve root symptoms, involve the instillation of an anaesthetic such as bupivacaine (Marcaine™) into the respective patho-anatomical structure suspected to be the source of LBP symptoms. Conducted by a suitably trained medical specialist, under aseptic conditions, with fluoroscopy guidance, FGFJI targets the facet joint, and TRNRB the intervertebral foramina. The premise is that when the symptomatic joint or innervating nerve is injected, temporary relief of

symptoms will result. If the joint injected is not the source of LBP, symptoms will be unaffected (Moult et al., 2013).

Clinically, these diagnostic procedures identify the symptomatic structure; then, using an 'intention to treat' protocol, a medical specialist may subsequently instil cortisone, such as methylprednisolone, using the same injection method. Thus, FGFJI and TFNRB are both diagnostic and treatment interventions that can provide immediate pain relief for patients with CLBP.

Placebo-controlled FGFJI is regarded the diagnostic 'gold standard', but clinically, financially and ethically, its use is controversial (Sehgal, Dunbar, Shah, & Colson, 2007). Consequently, although FGFJI that is not placebo controlled has not demonstrated high diagnostic accuracy (Curatolo & Bogduk, 2010), it represents a suitable reference standard for diagnostic studies (Bogduk, 2004). Similarly, TRNRB has demonstrated moderate diagnostic accuracy, is significant in surgical prognosis (Boswell et al., 2015) and is supported as a reference standard to diagnose radicular pain or nerve root symptoms (Datta et al., 2007).

Sacroiliac joint injection for the diagnosis of sacroiliac joint syndrome (SIJS) has shown moderate (Kennedy et al., 2015) and good diagnostic (Simopoulos et al., 2012) values, and therefore can be considered suitable as a reference standard. To diagnose discogenic pain, provocation discography is considered the gold standard (Stout, 2010). The diagnostic accuracy of other spinal injection procedures has not been widely researched, and is currently unsupported (Boswell et al., 2015).

1.3 Chronic Low Back Pain and Examination of Muscle Dysfunction

1.3.1 Theoretical mechanisms of functional loss

Regaining functional loss is a fundamental aim of physiotherapy in CLBP management. Nevertheless, scientific understanding of the mechanisms that underpin

functional loss in CLBP is currently lacking. There is evidence to suggest that reduced spinal motion (Christe, Redhead, Legrand, Jolles, & Favre, 2016), altered pain perception and intensity, and a complex interaction of motor control impairment in key lumbar stabilising muscles are factors (Dubois, Abboud, St-Pierre, Piché, & Descarreaux, 2014). In young adults with recurrent LBP, these factors appear related to the presence and interaction of reflex pain inhibition, muscle nerve supply deficit and reorganisation of the motor cortex, leading to alterations in corticomotor neural functions (Tsao, Danneels, & Hodges, 2011). Similar changes in motor cortex organisation have also been reported in some chronic musculoskeletal pain conditions, but as yet these changes are not well documented, and therefore unconfirmed in CLBP (Wand et al., 2011).

1.3.2 Motor control impairment

Motor control impairment, due in part to delayed activation of the transversus abdominis (TrA), has previously been reported in individuals with LBP (Hides et al., 2009). This muscle has received particular research focus because of its key role in lumbar spine stabilisation (Panjabi, 1992; Tesh, Shaw Dunn, & Evans, 1987). Lumbar segmental stability involves concomitant activation of several key stabilising muscles, with the TrA contributing via its attachment to the linea alba and thoraco-lumbar fascia's middle and posterior layers (Barker et al., 2006), and through its 'feed-forward' mechanism of activation (Rasouli, Arab, Amiri, & Jaberzadeh, 2011). Fascial tensioning due to TrA contraction results in transmission of tension across the lumbar spine, contributing to overall spinal stability (Barker, Briggs, & Bogeski, 2004).

Experimentally induced LBP has been shown to reduce TrA function (Hodges, Moseley, Gabrielsson, & Gandevia, 2003; Kiesel, Uhl, Underwood, &

Nitz, 2008). Alterations in lumbar postural control and stability, triggered by altered TrA activation (TrA-C), is a proposed mechanism for persistent LBP symptoms (Hodges & Moseley, 2003). Theoretically, perpetuation of CLBP may therefore be dependent on a recurring cycle of pain inhibition and TrA dysfunction.

Accordingly, assessment and treatment of TrA dysfunction are routinely conducted by physiotherapists to enable the patient to regain TrA function to reduce symptoms in CLBP (Richardson & Jull, 1995). Conversely, medical interventions such as DSAI that immediately abolish persistent symptoms in CLBP allow assessment and treatment of TrA dysfunction immediately following pain abolition.

Thus, immediate abolition of pain following DSAI provides physiotherapy researchers an opportunity to evaluate motor control impairment of the lumbar stabilising muscles in the absence of pain. An understanding of this may provide an original knowledge contribution about the immediate response of specific lumbar stabilising muscles whose dysfunction leads to or perpetuates functional loss in CLBP. Currently, the effect of immediate pain relief on spinal stabilising muscle activation and associated functional loss remains unreported in CLBP. The implications of pain relief specific to the immediate and direct effect on TrA function require investigation in CLBP.

1.3.3 Principles of real-time ultrasound imaging

Previous studies have reported various methodologies to assess TrA function (Hides et al., 2006; Hodges et al., 2003). One approach, real-time ultrasound imaging (RTUI), has been validated for thickness, length, volume and activation measurements of the TrA (Hides, Richardson, Jull, & Davies, 1995; Koppenhaver, Hebert, Fritz et al., 2009; McMeeken, Beith, Newham, Milligan, & Critchley, 2004). This form of ultrasound scanning images body structures in real-time using high

frequency sound waves (1–20 MHz; Lieu, 2010). The physical principles of RTUI relate to generation and transmission of these sound waves through the soft tissues and body fluids of the subject under examination.

An example of a machine can be seen in Plate 1.1. RTUI machines utilise the piezoelectric effect principle, whereby an electrical current is applied to a piezoelectric crystal enclosed within an ultrasound transducer or probe. The resultant short sound pulses are transmitted into the body and interact with anatomical structures to produce a two-dimensional image on the screen of the RTUI machine (Martin, Wells, & Goodwin, 2015). The sound waves can be reflected or transmitted, and this occurs in varying proportions, dependent on the type of tissue being scanned. The on-screen image (Plate 1.2) is a product of electrical impulses produced from sound waves as they are reflected back onto the piezoelectric crystal within the RTUI probe (Martin et al., 2015). Brightness of the on-screen images at tissue interfaces is directly related to differences in tissue density. Interfaces between anatomical structures with large differences in density will appear bright. An example of this is the soft tissue–bone interface, where a high percentage of ultrasound wave reflection is exhibited, indicating high acoustic impedance (Martin et al., 2015). The speed of the ultrasound is inversely proportional to the density of the structure it passes into (Lieu, 2010). Therefore, different tissues display variations in acoustic impedance. The same principle applies to the RTUI probe–air–skin interface. Air has almost 100% acoustic impedance, and necessitates the use of a water-based coupling gel between the probe and the patient's skin (Martin et al., 2015).

RTUI is a portable, inexpensive and non-invasive technology, which is suitable for use in a clinical setting. It can be used by suitably trained

physiotherapists to examine various musculoskeletal structures. Although safety factors such as the mechanical effect of micro-cavitation and a thermal effect from excessive heat production within the tissues should be considered, these are insignificant, provided the correct imaging settings are adhered to and length of routine scanning time is minimised (Martin et al., 2015).

1.3.4 Real-time ultrasound imaging to measure transversus abdominis activation

For more than 30 years, RTUI has been used to evaluate the spinal stabilising muscles and their function in both asymptomatic individuals and LBP patients (Gnat, Saulicz, & Miądowicz, 2012; Hides, 2006; Hides et al., 1995; Koppenhaver, Hebert, Fritz et al., 2009; Krag et al., 1987; McMeeken et al., 2004; Rostami et al., 2015; Saliba et al., 2010; Whittaker & Stokes, 2011). TrA-C, as an indicator of the muscle's functional ability, has been reported using RTUI TrA thickness measurement at rest, relative to that when fully contracted (Koppenhaver, Parent, Teyhen, Hebert, & Fritz, 2009; Lariviere et al., 2013), and provides clinicians with an empirical outcome measure. However, Whittle, Flavell and Gordon (2017) highlighted that most RTUI TrA studies focused on heterogeneous LBP study samples, with limited evidence and variable measurement reliability reported in CLBP.

1.3.5 Challenges to reliability of real-time ultrasound imaging in chronic low back pain

Whittle et al. (2017) identified methodological inconsistencies among RTUI TrA studies conducted in CLBP, and stated that reliability of RTUI TrA measurement reported in CLBP was sub-optimal. Hence, future research that

Plate 1.1. Real-time ultrasound machine.

assesses TrA function using RTUI in CLBP requires improved methods of image

acquisition and measurement.

Evidence suggests that sub-optimal TrA measurement reliability in CLBP

may be associated with anthropometric differences apparent in this population.

Specifically, increased body mass index (Cimolin et al., 2011; Heuch, Hagen, Heuch,

Nygaard, & Zwart, 2010), which decreases quality image capture (Brahee et al.,

2013; Ortiz, Chiu, & Fox, 2012). This is because excess subcutaneous fat distorts

ultrasound beam transmission (Miller, 2005), with potential for wider variation in

operator probe-to-skin pressure not evident in individuals with normal body mass
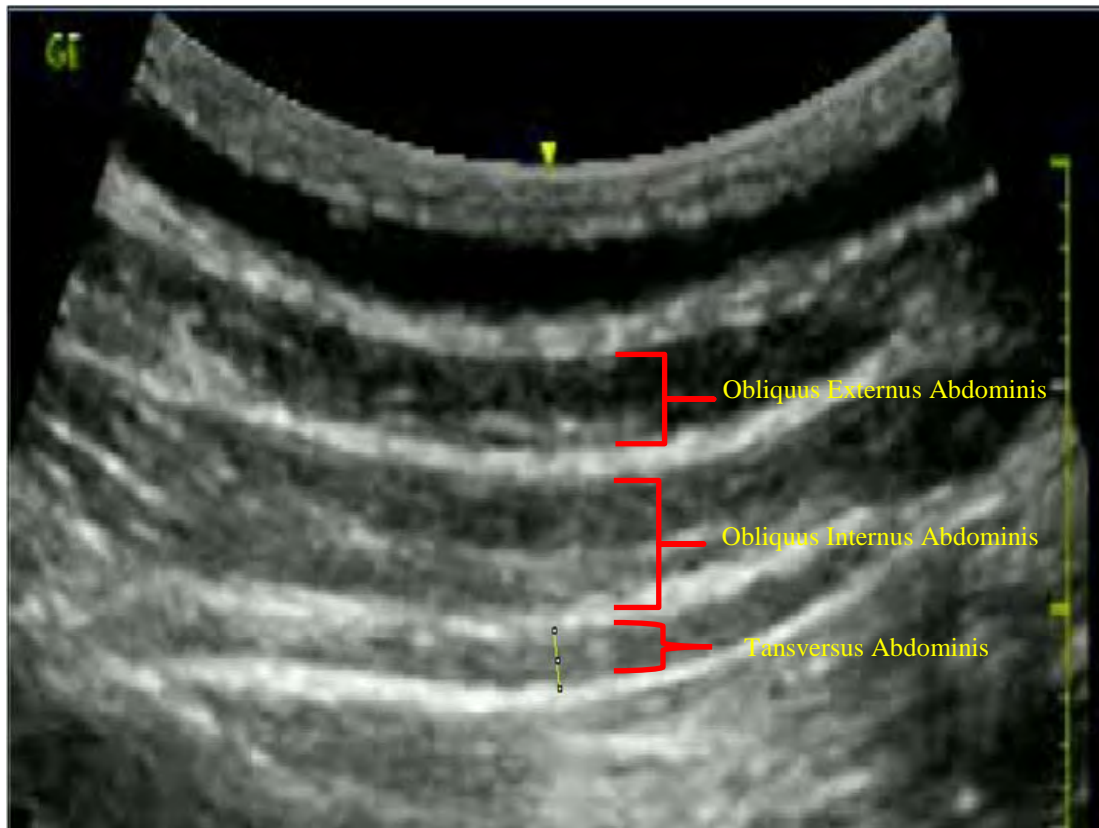
Plate 1.2. Real-time ultrasound on-screen image of obliquus externus abdominis, obliquus internus abdominis and transversus abdominis.

index (BMI). In addition, identification of bony landmarks becomes more challenging with increased BMI (Thanh Le, Robinson, & Lewis, 2015), which further compromises RTUI methodological standardisation. Whittle et al. (2017) identified that uncontrolled probe force, inclination and roll evident in 'free-hand' RTUI methods may in part explain the sub-optimal RTUI TrA measurement reliability reported in past CLBP research, and highlighted this of particular concern for researchers and clinicians.

Currently, RTUI machines have no in-built technology to limit variation in operator probe force, inclination and roll; thus, no standardised imaging methods to quantify and control for these factors are available. Left unaddressed, sub-optimal reliability of TrA RTUI prevails, particularly in challenging populations such as

those with high BMI. Therefore, technology specifically designed to reduce such confounding measurement factors in RTUI is required, and high intra-examiner reliability of such new methodology for TrA RTUI in CLBP must be established.

1.4 Statement of the Problem

Research specific to CLBP is required to improve and optimise treatment of this complex LBP subgroup. It is unknown whether currently used physiotherapy diagnostic classification processes for LBP are appropriate for CLBP which exhibits distinct confounding characteristics such as increased BMI and older age. In addition, little is known about the behaviour of the TrA following pain-relieving treatment interventions that abolish CLBP symptoms. Consequently, to improve physiotherapy management of CLBP, it is imperative that researchers strive to identify the most appropriate method of physiotherapy examination for these patients, and to understand the immediate effects of CLBP relief on the function of the TrA.

Efficacy of any physiotherapy examination process can only translate to the population under investigation. LBP is not a homogeneous group; thus, findings reported from studies conducted on heterogenic LBP populations should not be extrapolated to specific LBP subgroups. Accordingly, research related to physiotherapy examination of CLBP patients has clinical significance. Indeed, evidence suggests that the CLBP subgroup provides unique examination challenges for physiotherapists. Reported challenges include, but are not limited to, psychosocial influences, and demographic and physical characteristics. Which characteristic differences influence physiotherapy examination in these patients, and how they do so, is unclear.

Current evidence for a valid and reliable CLBP physiotherapy clinical examination pathway has not been critically appraised. It is therefore unclear what comprises an evidence-based examination algorithm specific to CLBP. Neither is it known whether the classifications used by physiotherapists to categorise CLBP clinical presentations agree with the diagnoses of a medical specialist in multidisciplinary CLBP management.

The physiotherapy profession appears to have adopted the assumption that results and recommendations from TrA RTUI studies conducted in mixed and acute LBP are appropriate to guide treatment interventions for CLBP. Whether this is an appropriate assumption is yet to be investigated. Further, the extent of motor control impairment in CLBP is not clearly understood. An evaluation of the immediate effect that CLBP abolition has on TrA function may contribute to this understanding. However, the validity and reliability of reports from such a study are dependent on optimal imaging, which is yet to be established.

1.5 Research Questions, Objectives and Hypotheses

1. **Question:** What physiotherapy lumbar spine examination processes or classification systems have demonstrated sufficient validity, diagnostic accuracy and reliability in CLBP for inclusion in a comprehensive physiotherapy CLBP examination algorithm?

    Statement of need: The validity and reliability of some physiotherapy classification systems and examination processes have been reported previously in heterogeneous samples of LBP patients. However, the validity and reliability when applied to CLBP patients are unclear.

Objective: To identify via systematic review, valid and reliable physiotherapy LBP classification systems and clinical examination tests with high diagnostic accuracy for CLBP populations.

2. **Question:** Are the physical measurements typically used by physiotherapists in LBP examination reliable to assess an asymptomatic population matched to, and typical of CLBP patients?

   Statement of need: Measurement of lumbar range of motion (ROM) and neuro-dynamic flexibility is a fundamental component of the CLBP examination. Factors such as obesity and older age associated with CLBP have potential to challenge the reliability of these measurements. No past study has reported such measurement reliability in a population matched for age and BMI typical of CLBP patients. Accordingly, it is unclear if these measurements are reliable for inclusion in a physiotherapy CLBP examination algorithm.

   Objective: To report intra-examiner reliability of lumbar spine and neuro-dynamic flexibility measurements in a simulated CLBP population.

   Hypothesis: It was hypothesised that, using standard protocols, a physiotherapist would measure lumbar spine ROM and neuro-dynamic flexibility reliably in a healthy population matched to the age and BMI characteristics of CLBP patients.

3. **Question:** What are the classification characteristics of a CLBP population using an evidence-based physiotherapy examination algorithm specifically designed for CLBP, including the demographics, Oswestry Disability Index (ODI), Roland–Morris disability questionnaire (RM),

Modified Somatic Pain Perceptions Questionnaire (MSPQ), symptom duration, and 11-point visual analogue scale (see Appendix 1)

Statement of need: Classification profile and characteristics of LBP populations have been reported in only two past studies (Eirikstoft & Kongsted, 2014; Hefford, 2008). Both included heterogeneous LBP study samples, with no descriptors of CLBP. The classification characteristics specific to a CLBP population have not been reported and remain unidentified.

Objective: On the basis of the outcomes of objectives 1 and 2, establish and apply a physiotherapy CLBP movement and patho-anatomical-based examination algorithm, along with age, gender, BMI, RM and ODI disability scores, visual analogue scale (VAS) pain intensity, pain distribution and MSPQ pain somatisation scores, to report the classification characteristics of a CLBP population.

Hypothesis: It was hypothesised that, because of the association of older age and higher BMI in CLBP, physiotherapy classifications aligned with degenerative facet joints and intervertebral discs would predominate, and that the biopsychosocial complexity reported in CLBP would result in moderate to high reported levels of disability, pain intensity and somatisation.

4. **Question:** How well does the diagnostic classification of a physiotherapist using a comprehensive CLBP examination algorithm agree with diagnostic reference standards?

Statement of need: Multidisciplinary assessment is important for optimal management of CLBP. However, this requires effective interdisciplinary

communication and shared diagnostic language. Previously, two within-discipline agreement studies have been conducted. However, interdisciplinary levels of diagnostic agreement have not been reported. In particular, CLBP diagnostic classification agreement between a physiotherapist and a neurosurgeon specialised in spinal examination has yet to be reported. Further, the level of combined clinical diagnostic agreement of both examiners relative to an available reference standard is currently unknown.

Objective: In a CLBP population, determine the diagnostic agreement between the clinical classification by an experienced physiotherapist, using a comprehensive CLBP examination algorithm, and a reference standard, comprising a CLBP specialised neurosurgeon's diagnosis derived from physical examination, MRI or DSAI, as clinically indicated. In addition, report the combined clinical diagnostic agreement of both examiners relative to the outcome of a reference standard DSAI to diagnose FJS and FJS-related nerve root compromise.

Hypothesis: It was hypothesised that a high level of agreement would be demonstrated between the physiotherapist's classification and the reference standard diagnosis, and the combined clinical diagnostic agreement of both examiners relative to the outcome of DSAI.

5. **Question:** In a CLBP study sample, what is the utility of a novel equipment developed to address TrA RTUI acquisition confounders of uncontrolled probe force, inclination and roll, and is TrA RTUI measurement of a single physiotherapist using the novel equipment reliable?

Statement of need: Inconsistent methodologies and variable reliability of RTUI TrA measurement specific to CLBP have been identified (Whittle et al., 2017). Confounding factors that compromise RTUI TrA measurement reliability relate to increased BMI, which predominates in CLBP. A method that controls and standardises RTUI probe pressure, inclination and roll may limit confounders of TrA measurement in CLBP. To date, no method to control these RTUI probe parameters has been established, and this requires further investigation.

Objective: To trial the utility and intra-examiner reliability of a novel RTUI TrA image acquisition tool that standardises probe force, inclination and roll.

Hypothesis: It was hypothesised that use of the novel equipment would achieve high intra-examiner reliability of RTUI TrA measurement, superior to that which has been reported previously in CLBP populations.

6. **Question:** What is the immediate effect of pain relief on TrA-C in a CLBP population?

Statement of need: The deleterious effect of experimentally induced LBP on TrA-C has been reported previously. However, there remains no understanding of the immediate effect of CLBP abolition on TrA-C, which requires further investigation.

Objective: To assess TrA-C, using reliable RTUI methodology, before and immediately after pain abolition achieved from the anaesthetising effect of DSAI in a CLBP population.

Hypothesis: Due to the multi-factorial presentation of CLBP, it was hypothesised that immediate pain relief from DSAI would result in no statistical or clinically significant difference in pre–post TrA-C.

1.6 Thesis Structure

There were three primary foci of this research, which are addressed in distinct yet interconnected stages. Each chapter of this thesis addresses a specific research objective, includes the associated published manuscript or the study in manuscript format (under review), and concludes with a summary of key findings and how these directed the progression of the research.

This research used a stage-based process whereby one or more preceding studies guided the following stage of the overall project. The chapters of this thesis are presented in the same sequential manner. The rationale behind this process was born from the fundamental prerequisite that each clinical study would be founded on the most valid and reliable methodologies. Portable document format versions of published articles ($N = 3$) associated with this research are provided at the end of the thesis as appendices.

1.6.1 Preparation

This section reports the literature review and preparatory projects that underpinned the subsequent studies. Specifically, systematic reviews of diagnostic accuracy and reliability of CLBP clinical examination processes used by physiotherapists. These studies guided which examination components were most appropriate to include in a comprehensive physiotherapy examination algorithm for CLBP. This part of the thesis also reports results of a subsequent study, conducted to establish intra-examiner reliability for the lumbar spine clinical measurements chosen for inclusion in the physiotherapy examination algorithm for CLBP.

1.6.2 Application of a physiotherapy chronic low back pain examination algorithm

This section reports two studies that applied a comprehensive physiotherapy examination algorithm for CLBP classification. One reported CLBP characteristics and classification category frequencies, and compared CLBP characteristics among classification categories. The other reported classification agreement between a physiotherapist's clinical examination using the algorithm and available reference standards.

1.6.3 Measurement of the transversus abdominis using real-time ultrasound

This section presents a novel new equipment using technology developed to quantify and standardise RTUI probe parameters of force, inclination and roll. This provided an opportunity to implement an updated method for TrA RTUI in CLBP patients, and report a study conducted to evaluate intra-observer RTUI TrA measurement reliability resulting from the use of the novel new equipment.

This part of the thesis includes the final study of this research. This study reported the immediate effect of CLBP abolition following DSAI on TrA-C measured using the novel new equipment to standardise force, inclination and roll across measurements during RTUI.

1.7 Research Context

All studies were conducted in Townsville, a regional city in Northeast Queensland, Australia. Study sites included the physiotherapy practical teaching rooms of James Cook University; the Townsville Hospital Persistent Pain Management Clinic, Douglas; and the Mater Hospital Spinal Injection Clinic, Pimlico.

Townsville Hospital is a tertiary public hospital servicing over 230,000 people, north to the Cape York Peninsula and Torres Strait Islands, and west to Mount Isa and the Gulf of Carpentaria. The Mater Hospital is a private hospital with 150 beds situated in Townsville city, and is part of Mater Health Services, North Queensland. Both Townsville Hospital and Mater Hospital are teaching hospitals for the James Cook University health programmes.

Chapter 2: Diagnostic Accuracy of Physiotherapy Examination and Classification in Chronic Low Back Pain

This chapter in combination with Chapters 3 and 4 addresses the first research question (see Chapter 1): 'What physiotherapy lumbar spine examination processes or classification systems have demonstrated sufficient validity, diagnostic accuracy and reliability in CLBP for inclusion in a comprehensive physiotherapy CLBP examination algorithm?'

2.1 Introduction

Of Greek derivation, diagnosis is 'to know through'. Alternatively:

In the context of medicine, it is to see through the patient's symptoms and other findings to imagine and understand what may be happening in terms of current theories applied to medicine. The decision of what to do is made by using the diagnosis to infer what will probably happen next and how the process can be changed by various available interventions. (Lipschik, Von Feldt, & Frame, 2009, p. xii)

The diagnostic process that physiotherapists and other health practitioners use is often ambiguous, and depends on a series of deductive actions and past experience to rule out other potential diagnoses to reach a final diagnosis. Although clinical diagnosis and classification have slightly different definitions, they may be used synonymously to indicate categorisation or labelling of a patient's condition (Kutschenko, 2011; Lipschik et al., 2009).

Correct diagnostic classification is reliant on the accuracy of the classification systems, examination processes, tests or clinical prediction rules (CPRs) that are conducted prior to the final diagnostic decision. Clinical prediction rules are derived from research which clinical presentations optimally predict the probability that a

structure or condition is present (Laupacis, Sekar,& Stiell, 1997). The accuracy of classification systems, examination processes, tests or CPRs to measure or evaluate what they are actually developed for is defined as measurement validity (Portney & Watkins, 2009). Simply stated, how well do they do what they purport to do, and what is the clinician able to do in response to the results they provide?

Although there are several types of measurement validity, the most objective is criterion-related validity (Portney & Watkins, 2009). One subdivision of criterion-related validity is concurrent validity, which compares the ability of a test (index test) to predict the outcome or response of another highly valid test (gold or reference standard) in the same group of participants. The gold standard must be highly validated and therefore accurate. If a gold standard (100% accurate) is not available, then a suitable reference standard may be used, even though its accuracy is known to be imperfect (Rutjes, Reitsma, Coomarasamy, Khan, & Bossuyt, 2007).

In criterion-related concurrent diagnostic accuracy or validity studies, the response to the test or examination process is most often dichotomous, either 'present' or 'absent'. The data obtained are categorical, and the statistical analysis usually includes chi-square analysis, reporting sensitivity (Sn), specificity (Sp), likelihood ratios (LRs) and predictive values, with regression analysis or receiver operating characteristic (ROC) analysis. The index test is considered highly accurate when its outcome correlates highly with the gold or reference standard with which it is compared.

Ford, Story, O'Sullivan and McMeeken (2007) conducted a systematic review of the methods used to develop and validate LBP classification processes, and concluded that methods and validation were unclear and should be a research priority. Those authors also suggested that development of classification processes

should examine all aspects of LBP, such as patho-anatomical, psychosocial and neurophysiological factors, as failure to do so may result in an incomplete evaluation of this complex condition (Ford et al., 2007).

Previously, the diagnostic accuracy of clinical examination in LBP patients has been the subject of several systematic reviews (Al Nezari, Schneiders, & Hendrick, 2013; Alqarni, Schneiders, & Hendrick, 2011; Ford et al., 2007; Hancock et al., 2007; Littlewood & May, 2007; May & Aina, 2012; Murphy, Hurwitz, & Nelson, 2008; Petersen, Laslett, & Juhl, 2017; Petersen, Thorsen, Manniche, & Ekdahl, 1999; Scaia, Baxter, & Cook, 2012; Sivayogam & Banerjee, 2011); most of those conducted in the past 10 years have reported the diagnostic accuracy of individual or clustered tests for biological sources of LBP (Al Nezari et al., 2013; Alqarni et al., 2011; Hancock et al., 2007; Littlewood & May, 2007; May & Aina, 2012; Murphy et al., 2008; Petersen et al., 2017; Scaia et al., 2012; Sivayogam & Banerjee, 2011). The diagnostic accuracy reported from these reviews has been variable, and no studies appear to report diagnostic classification validity of comprehensive physiotherapy classification systems applied exclusively in CLBP study samples.

The inclusion of, and recommendations for using patho-anatomical diagnosis as part of physiotherapy examination and classification is a contentious issue within the profession. Particularly considering that clinical guidelines do not support a purely patho-anatomical approach. Indeed, Henschke (2006) suggested that guidelines often rely on previous guidelines, narrative review and do not always consider original research. Van Zundert et al. (2013) stated that guidelines sometimes provide contradictory recommendations of which one perspective may be used by funding bodies and policy makers for decisions on financing treatment. They

suggested that this potentially challenges the evidential support for guidelines, and suggested that for LBP the diagnosis should where possible investigate the most likely structural source and level of symptoms prior to treatment.

Hence, two schools of thought exist which appear to be somewhat polarised. Recently, both perspectives have been debated (Saragiotto et al., 2017). Regardless of the chosen perspective, the limited and variable validity of CLP examination systems, tests and CPRs currently reported, predominantly relates to heterogeneous LBP populations. Consequently, the summative conclusions from these reviews cannot be extrapolated directly to CLBP.

The objective of this review was to appraise published literature that has reported validity or diagnostic accuracy of physiotherapy LBP classification systems, physical examination tests, test clusters or CPRs in CLBP-specific populations, and then to use these findings to propose valid examination methods for inclusion in a comprehensive physiotherapy CLBP examination algorithm.

2.2 Method

A systematic review of primary research that involved location, appraisal and synthesis of published studies was conducted. This study design constituted Level 1 evidence according to the National Health and Research Council level of evidence (National Health and Medical Research Council, 2009).

2.2.1 Search strategy

The following databases were accessed: Medline via OvidSP (1946 to December 2016); CINAHL (no date restriction); PEDro (no date restriction); the Cochrane library (no date restriction); and Informit (1970 to December 2016). A keyword, title and abstract search was conducted using specific search terms, truncated as required (see Table 2.1). Between search terms, the Boolean operators

'and' and 'or' were used. The database search was initially conducted on 9 September 2013, and updated on 20 December 2016.

2.2.2 Study selection

Article titles were screened and, if appropriate, were imported into EndNote (version 17.3.1). The abstracts and reference lists of the imported publications were scanned to identify articles suitable for full text inclusion. The identified full text articles were appraised by the reviewer for inclusion according to the following criteria.

2.2.3 Eligibility criteria

2.2.3.1 Types of studies

Studies that reported the validity or diagnostic accuracy of complete physiotherapy LBP classification systems, individual physical examination tests, test clusters or CPRs were included.

2.2.3.2 Participants

All study participants were required to be adult ($\geq$18 years old) and to have been experiencing LBP, with or without associated lower limb symptoms, continuously for 12 weeks or more.

2.2.3.3 Type of outcome measure or intervention

Studies were eligible for inclusion if they reported either agreement due to chance via kappa coefficients, Sn, Sp, predictive values or any statistical analysis appropriate to establish concurrent validity or diagnostic accuracy, or diagnostic CPRs.

Table 2.1

Search Strategy

1	Low back pain*	AND	Validity* or Accuracy*	AND	Classification* or Diagno* or Examination* or Screening*
2	Low back pain*	AND	Validity* or Accuracy*	AND	Mechanical diagnosis and treatment* or Treatment-based classification* or Patho-anatomical* or Movement system impairment classification* or O'Sullivan classification system* or Motor control impairment* or McKenzie assessment*
3	Clinical prediction rule* or Clinical prediction* or Predictive value of tests*	AND	Lumbar vertebrae* or Lumbar*	AND	Facet* or Zygapophyseal joint* or Intervertebral disc degeneration* or Discogenic* or Intervertebral disc* or Instability* or Sacroiliac joint*

2.2.3.4 Participants

All study participants were required to be adult ($\geq$18 years old) and to have been experiencing LBP, with or without associated lower limb symptoms, continuously for 12 weeks or more.

2.2.3.5 Type of outcome measure or intervention

Studies were eligible for inclusion if they reported either agreement due to chance via kappa coefficients, Sn, Sp, predictive values or any statistical analysis appropriate to establish concurrent validity or diagnostic accuracy, or diagnostic CPRs.

2.2.4 Exclusion criteria

Systematic reviews, unpublished studies, studies from non-peer-reviewed publications, opinion, consensus or discussion papers, and those not published in English, were excluded from this study. Also excluded were studies that reported results from classification systems not routinely used by physiotherapists, examination based only on clinical observation, self-reported questionnaires and other non-physical examination methods. Further exclusions were studies with participants whose symptoms were less than 12 weeks' duration (acute and subacute LBP); who were pre- or post-partum; who were diagnosed with inflammatory disease, malignancy or pain of visceral origin; or who were post spinal surgery.

2.2.5 Data extraction, synthesis and analysis of results

Each eligible full text article was appraised using the quality assessment of comparative diagnostic accuracy studies (QUADAS)-2 (Wade, Corbett, & Eastwood, 2013; Whiting et al., 2011). This is a valid tool to appraise diagnostic studies, and uses 'signalling questions' to assess the 'risk of bias' and 'concerns regarding applicability' for patient selection, index tests, reference standards and flow of

patients through the study. Articles were scored as 'yes', 'no' or 'unclear'. If each

signalling question resulted in a 'yes' response, a low 'risk of bias' or 'concern

regarding applicability' was indicated. A 'no' response recorded for any signalling

questions indicated potential 'risk of bias' or 'concern regarding applicability'. The

reviewer scored the final full text articles according to guidelines stated in the

QUADAS background document (see Appendix 2).

2.3 Results

2.3.1 Study selection

The search strategy (see Table 2.1) identified 1239 research articles from the

electronic databases. These were screened and reviewed (see Figure 2.1), according

to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses

guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). Following removal of

duplicate articles, title screening and inclusion of articles identified from reference

list scanning, 71 articles remained. Abstract review was conducted, and 17 articles

were eligible for full text screening.

Full text screening resulted in rejection of 12 articles (Abbott et al., 2006;

Abbott et al., 2005; Ahn & Jhun, 2015; Fritz, Piva, & Childs, 2005; Laslett, Aprill,

McDonald, & Young, 2005; Laslett & van Wijmen, 1999; Laslett, Young, Aprill, &

McDonald, 2003; Paul et al., 2008; Petersen et al., 2003; Stojanovic & Engel, 2015;

Tousignant, Poulin, Marchand, Viau, & Place, 2005; Werneke et al., 2011). Reasons

for rejection included studies ($N = 1$) that did not report validity, diagnostic accuracy

or diagnostic CPRs (Abbott et al., 2006) and studies ($N = 11$) in which the study

population was not exclusively CLBP patients (Abbott et al., 2005; Ahn & Jhun,

2015; Fritz et al., 2005; Laslett, Aprill, et al., 2005; Laslett & van Wijmen, 1999;

```
┌─────────────────────────────┐
│ Total number of articles    │
│ identified from database    │
│ searches = 1239             │
└─────────────────────────────┘
              │
              │          ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
              │            Excluded:
              │          │   Duplicates and title screen =1168 │
              ◄──────────   Included:
              │          │   Identified articles from scanned  │
              │            reference lists = 0
              │          └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
              ▼
┌─────────────────────────────┐
│ Total number of abstracts   │
│ screened = 71               │
└─────────────────────────────┘
              │
              │          ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
              │            Abstracts excluded:
              │          │   Not relevant or eligible = 18 │
              ◄──────────   Not diagnostic studies = 4
              │          │   Not CLBP = 15                 │
              │          └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
              ▼
┌─────────────────────────────┐
│ Total number of full text   │
│ articles screened = 17      │
└─────────────────────────────┘
              │
              │          ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
              │            Full text articles excluded:
              │          │   Not exclusively CLBP = 11     │
              ◄──────────   Not diagnostic studies = 1
              │          └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
              ▼
┌─────────────────────────────┐
│ Total number of studies appraised for │
│ 'applicability' and 'risk of bias' = 5 │
└─────────────────────────────┘
```

Figure 2.1. Flow chart of study selection process.

Laslett et al., 2003; Paul et al., 2008; Petersen et al., 2003; Stojanovic & Engel, 2015; Tousignant et al., 2005; Werneke et al., 2011).

Five eligible full text articles were accepted for final appraisal, and key study details were summarised (see Table 2.2). Accuracy to diagnose or predict the presence of discogenic pain was reported in two studies (Laslett, Aprill, McDonald, & Öberg, 2006; Laslett, Öberg, Aprill, & McDonald, 2005), of zygapophyseal pathology (hereafter referred to as FJS) in two studies (Laslett, McDonald, Aprill, Tropp, & Öberg, 2006; Laslett, Öberg, Aprill, & McDonald, 2004) and of chronic radiculopathy in one study (Iversen et al., 2013).

2.3.2 Risk of bias within studies

The QUADAS-2 appraisal tool was applied to each of the five studies. Overall evaluation for risk of bias across all studies using the QUADAS-2 is shown in Figure 2.2. 'Flow and Timing' was the domain with the highest risk of bias (100%). Risk of bias was also evident in the domains of 'Index Test' (60%) and 'Patient Selection' (40%). There was no risk of bias for the reference standard domain.

2.3.3 Applicability judgement within studies

Overall concern regarding applicability of studies following appraisal with the QUADAS-2 is shown in Figure 2.2. Across all three domains of 'Index Test', 'Patient Selection', and reference standard there were no concerns regarding study applicability for this review.

2.3.4 Summary of results

2.3.4.1 Predictors of chronic radiculopathy

Iverson et al. (2013) investigated the accuracy of clinical examination in CLBP patients with radiculopathy, referred to specialist care, to identify nerve root

Table 2.2

Summary Description and Results of Included Articles

Study	Objective	Index test/s	Reference standard	Method	Statistical analysis	Summary of findings
Iversen et al. (2013)	To evaluate the accuracy of clinical index tests to identify lumbar nerve root impingement in patients with chronic radicular pain in a specialised care setting.	SLR Femoral nerve stretch. Lower limb myotome and dermatome tests. Tendon reflex testing.	MRI (94.0%) or CT (6.0%)	Consecutive recruitment. Examiner blinded to reference standard. Neuro-radiologist blinded to index test diagnosis.	Sn, Sp, and LR, including 95% CI, were calculated for each clinical test. ROC curve and an estimate for the AUC were reported.	Accuracy of individual index tests was low. Accuracy of tests combined was slightly higher.

Note. AUC = area under the curve; CAT = computerised axial tomography; CI = confidence interval; CLBP = chronic low back pain; CP = centralisation phenomenon; CPR = clinical prediction rule; LR = likelihood ratio; MRI = magnetic resonance imaging; ROC = receiver operating characteristic; Sn = sensitivity; Sp = specificity; SLR = straight leg raise test; ZJ = zygapophyseal joint.

Table 2.2

Summary Description and Results of Included Articles (Continued)

Study	Objective	Index test/s	Reference standard	Method	Statistical analysis	Summary of findings
Laslett, Aprill, et al. (2006)	To estimate the diagnostic accuracy of clinical variables in relation to provocation discography for pain-sensitive discs.	A history and structured physical examination included: - a McKenzie-styled assessment. - range of motion tests - neuro-dynamic tests - sacroiliac joint and facet joint provocation tests - standardised repeated end-range movements - presence of vulnerability in neutral zone.	Discography	Examiner blinded to reference standard. Discographer blinded to index test diagnosis.	Sn, Sp and LR calculated including CI for each test Pre- and post-test odds and post-test probabilities were calculated. Chi-square, Fisher's exact test and univariate and multivariate regression analysis conducted for independence of examination outcome variables.	Signs of possible discogenic pain with positive discography may be identified in CLBP using CP and combinations of: - loss of lumbar extension or - reported vulnerability

Note. AUC = area under the curve; CAT = computerised axial tomography; CI = confidence interval; CLBP = chronic low back pain; CP = centralisation phenomenon; CPR = clinical prediction rule; LR = likelihood ratio; MRI = magnetic resonance imaging; ROC = receiver operating characteristic; Sn = sensitivity; Sp = specificity; SLR = straight leg raise test; ZJ = zygapophyseal joint.

Table 2.2

Summary Description and Results of Included Articles (Continued)

Study	Objective	Index test/s	Reference standard	Method	Statistical analysis	Summary of findings
Laslett, McDonald, et al. (2006)	To estimate the predictive power of clinical findings in relation to pain reduction after screening ZJ blocks.	A history and structured physical examination included: - a McKenzie-styled assessment - range of motion.tests - neurological tests - neuro-dynamic tests - sacroiliac joint and facet joint provocation tests. - standardised repeated end-range movements.	Intra-articular ZJ joint injection or medial branch block.	Examiner blinded to reference standard. Interventional radiologist blinded to index test diagnosis.	Sn, Sp and LR calculated including CI for each test. ROC curve and an estimate for the AUC were reported.	Tests showed value to rule out a positive response to a ZJ block when: - CP was present or CPR5 was negative (3 or more of 5 clinical signs: - age >50 - symptoms best walking or sitting - paraspinal pain - +ve extension/rotation test).

Note. AUC = area under the curve; CAT = computerised axial tomography; CI = confidence interval; CLBP = chronic low back pain; CP = centralisation phenomenon; CPR = clinical prediction rule; LR = likelihood ratio; MRI = magnetic resonance imaging; ROC = receiver operating characteristic; Sn = sensitivity; Sp = specificity; SLR = straight leg raise test; ZJ = zygapophyseal joint.

Table 2.2

Summary Description and Results of Included Articles (Continued)

Study	Objective	Index test/s	Reference standard	Method	Statistical analysis	Summary of findings
Laslett, Öberg, et al. (2005)	To estimate predictive power of centralisation and the influence of disability and patient distress on diagnostic performance, using provocation discography.	A history and structured physical examination included: - a McKenzie-styled assessment - range of motion tests - neuro-dynamic tests. - sacroiliac joint and facet joint provocation tests - standardised repeated end-range movements.	Discography	Examiner blinded to reference standard. Discographer blinded to index test diagnosis.	Sn, Sp and LR calculated, including CI, for each test. Multiple logistic regression analysis conducted to identify variables influential in the prediction of discography outcome.	The report of CP observed using the McKenzie assessment in non-distressed and not severely disabled (Sp of 89%), distressed and severely disabled (Sp of 100%) CLBP patients indicated that, during an initial McKenzie assessment in patients without severe disability or distress, positive discography and a diagnosis of discogenic pain is highly likely. This high Sp enables clinicians to reason a discogenic pain diagnosis

Note. AUC = area under the curve; CAT = computerised axial tomography; CI = confidence interval; CLBP = chronic low back pain; CP = centralisation phenomenon; CPR = clinical prediction rule; LR = likelihood ratio; MRI = magnetic resonance imaging; ROC = receiver operating characteristic; Sn = sensitivity; Sp = specificity; SLR = straight leg raise test; ZJ = zygapophyseal joint.

Table 2.2

Summary Description and Results of Included Articles (Continued)

Study	Objective	Index test/s	Reference standard	Method	Statistical analysis	Summary of findings
Laslett et al. (2004)	To evaluate the utility of 'Revel's criteria' as a screening tool for selection of CLBP patients for controlled ZJ diagnostic blocks.	A history and structured physical examination included: - a McKenzie-styled assessment. - range of motion tests - neuro-dynamic tests - neurological tests - sacroiliac joint and facet joint provocation tests - standardised repeated end-range movements - Revel's criteria (standing flexion, returning from standing flexion, standing extension, the extension rotation test).	Intra-articular ZJ joint injection or medial branch block.	Examiner blinded to reference standard. Interventional radiologist blinded to index test diagnosis.	Sn, Sp and LR calculated, including CI, for each test.	Revel's criteria were not suitable (low Sn and high Sp) as a clinical screening tool to select which CLBP patients are suitable for diagnostic ZJ blocks. The criteria cannot be considered diagnostic of symptomatic lumbar ZJ.

Note. AUC = area under the curve; CAT = computerised axial tomography; CI = confidence interval; CLBP = chronic low back pain; CP = centralisation phenomenon; CPR = clinical prediction rule; LR = likelihood ratio; MRI = magnetic resonance imaging; ROC = receiver operating characteristic; Sn = sensitivity; Sp = specificity; SLR = straight leg raise test; ZJ = zygapophyseal joint.

(NR) impingement from intervertebral disc herniation. In cooperation with a physiotherapist, and following prior training, neurology, rehabilitation, and physical medicine specialists conducted clinical examinations using four index tests (see Table 2.2).

Neuro-radiologists examined MRI or computerised tomography (CT) images, and, as reference standards, reported the presence or absence of discogenic NR impingement based on the imaging results. The number of clinical and imaging examiners was not specifically stated. Results indicated that no individual test had high accuracy in CLBP. Sn and Sp were low with wide confidence intervals (CIs), positive LRs (+ve LRs) were ≤ 4.0 and all negative LRs ($-$ve LRs) were ≥ 0.4. However, combining tests showed a slight improvement in accuracy to predict NR impingement. The ROC analysis resulted in an area under the curve of .95 (95% CI [0.90, 1.00]) for fourth lumbar vertebral level (L4); .67 (95% CI [0.56, 0.77]) for fifth lumbar vertebral level (L5); and .66 (95% CI [0.54, 0.77]) for sacral vertebral level one (S1) NR impingement.

It was concluded that use of the straight leg raise test (SLR), the femoral nerve stretch, dermatome and myotome testing, or tendon reflex clinical tests does not demonstrate sufficient accuracy to predict whether NR impingement is present or absent when applied separately. However, there was improved accuracy to predict the presence of NR impingement in patients with chronic radicular pain in a specialist care setting if the tests were conducted together as a test cluster.

This study had the lowest overall risk of bias, compared with the other four studies in this review. Only the 'Flow and Timing' domain of the QUADAS-2 indicated any risk of bias. All patients received the reference standard and were included in the analysis; however, bias might have been introduced because the
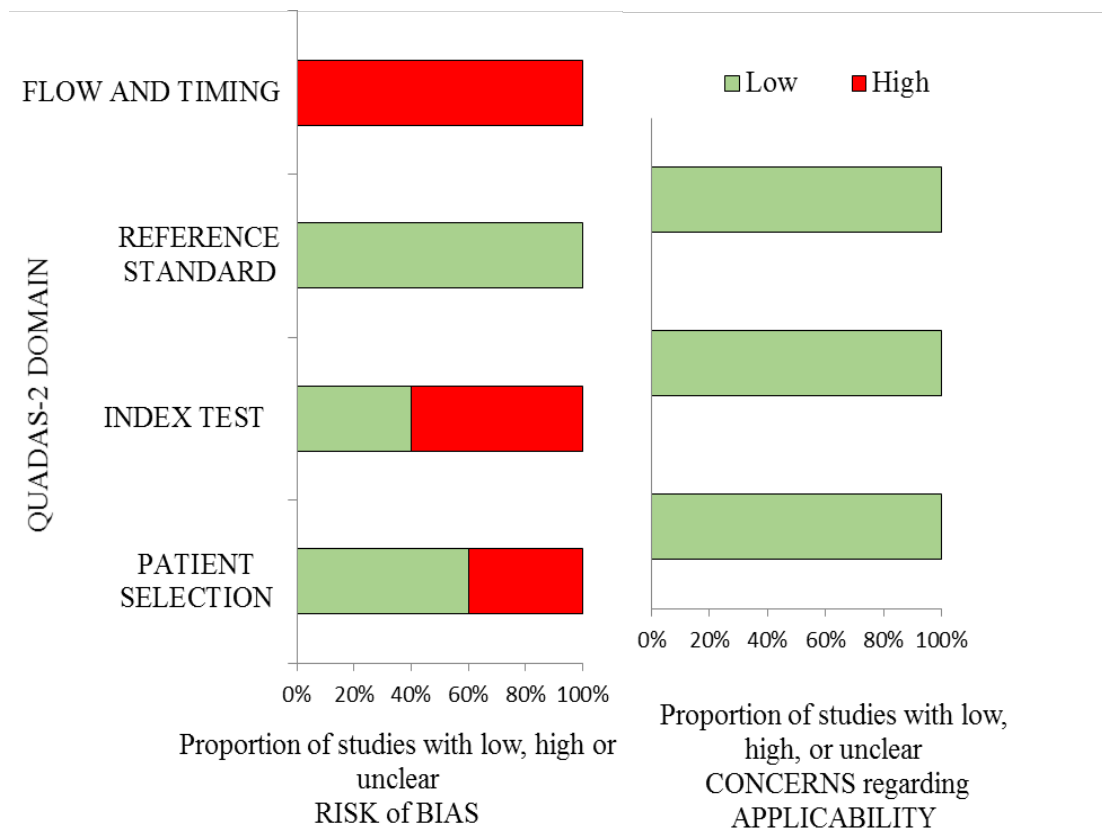
Figure 2.2. Risk of bias across all studies using the quality assessment of comparative diagnostic accuracy studies-2.

interval between the index tests and the reference standard was not stated, and participants did not all receive the same reference standard (MRI 94%; CT 6%).

2.3.4.2 *Predictors of facet joint syndrome*

Laslett et al. (2004) investigated the accuracy of Revel's criteria (Revel et al., 1998) as a screening test for FJS in patients referred to a diagnostic radiology clinic. Clinical examination was conducted by two physiotherapists using an examination synonymous with the McKenzie Institute Lumbar Spine Assessment (McKenzie Institute International [MII], 2005).

The McKenzie Institute Lumbar Spine Assessment (MII, 2005) is embedded in contemporary physiotherapy practice and utilises movement-based processes to establish baseline range and quality of movement, followed by repeated movement

testing (McKenzie, 1981; McKenzie & May, 2003) to identify the presence or absence of directional preference (DP) alone or DP plus centralisation phenomenon (CP). This movement-based system classifies patients with spinal symptoms into three main syndromes—derangement, dysfunction and postural (McKenzie & May, 2003). Patients who do not fit the three main syndromes are classified as 'other'. Laslett et al. (2004) reasoned the presence or absence of Revel's criteria (Revel et al., 1998) from components of the McKenzie examination. A radiologist examined patients, reviewed imaging and, as clinically indicated, conducted the reference standard diagnostic injections relative to identified symptomatic structures. Two facet joint–related diagnostic injection procedures, FGFJI or medial branch block, were conducted by one of two 'injectionists'.

Two different pain intensity reduction cut-offs were applied in this study. Reference standard A, a reduction of greater than 75% in pre–post injection pain intensity reported on a VAS, to indicate a positive diagnosis of FJS ($N = 108$), replicated that of Revel et al. (1998). Reference standard B required complete abolition of the patient's pain ($N = 18$). Results indicated that logistic regression was not significant for reference standard A ($N = 108$, $p = .46$) or B ($N = 100$, $p = .06$); therefore, Revel's model is not sufficiently accurate to select CLBP patients suitable for facet joint injection. Sn was low and Sp high, +ve LRs were ≤ 4.0 and −ve LRs ≥ 0.4 for both reference standards. It was concluded that the use of Revel's criteria (Revel et al., 1998) does not predict the presence or absence of FJS in patients with CLBP referred to a radiology clinic.

In this study, design, patient exclusion and blinding were appropriate, and all patients received the same reference standard. However, an overall risk of bias was demonstrated across three of the QUADAS-2 domains. 'Patient Selection' might

have been subject to bias because it was unclear if the study participants were enrolled consecutively or randomly. 'Index Test' might have been subject to bias because no pre-specified index test threshold was stated. 'Flow and Timing' was subject to the highest risk of bias because not all patients were included in the analysis, and the interval between the index tests and the reference standard was not stated. The reference standard domain of the QUADAS-2 indicated no risk of bias.

Clinical predictors for FJS were also investigated by Laslett, McDonald, et al. (2006) in patients referred to a radiology clinic, using a clinical examination synonymous with the McKenzie Institute Lumbar Spine Assessment (MII, 2005), plus additional patho-anatomical provocation tests. Examiners were two physiotherapists credentialed in the McKenzie method of examination (McKenzie & May, 2003). The reference standard was conducted according to methods reported by Laslett et al. (2004). To indicate a positive diagnosis of FJS, five different pain intensity reduction cut-offs were applied in this study, above 75%, 80%, 85%, 90% and 95%.

Results indicated that sufficient diagnostic accuracy was reached at the 90% and 95% reduction in pain intensity levels, and this was consistent across all variables at the 95% level. The absence of CP and a positive extension rotation test (ER) showed 100% Sn, but low Sp. The variable of 'no pain on rising from sitting' showed no diagnostic utility for FJS. However, a negative test response to the variables of age over 50 years, pain best when walking or sitting, paraspinal symptom onset, and the ER test, combined with a pain somatisation score less than 13 on the MSPQ (Main, 1983), in the presence of CP, showed value to rule out a positive response to facet joint block. Thus, in combination these variables indicate diagnostic utility to rule out the presence of FJS in patients with CLBP.

In this study, design, patient exclusion, blinding, and time interval between index test and reference standard were appropriate, and all patients received the same reference standard. However, this study demonstrated an overall risk of bias across three of the QUADAS-2 domains. 'Patient Selection' might have been subject to bias because it was unclear if the study participants were enrolled consecutively or randomly. 'Index Test' might have been subject to bias because no pre-specified index test threshold was stated. 'Flow and Timing' might have been subject to bias because not all patients received the reference standard, and not all were included in the analysis. The reference standard domain of the QUADAS-2 indicated no risk of bias.

2.3.4.3 Predictors of discogenic pain

Laslett, Öberg, et al. (2005) investigated the predictive power of CP using provocation discography in CLBP. An experienced manipulative physiotherapist and former senior instructor at the MII conducted all clinical examinations, which were synonymous with the McKenzie Institute Lumbar Spine Assessment (MII, 2005), with additional patho-anatomical provocation tests and an evaluation of disability, and distress represented as a function of depression, pain intensity and somatisation scores (see Table 2.2). Two discographers conducted provocative discography as the reference standard to identify the presence or absence of discogenic pain.

Results indicated that Sn was low, but Sp was high for CP as a predictor of a positive discograph (+ve LR of 6.9, and −ve LR of 0.63), but all CIs were wide. Multiple regression analysis indicated that high disability, pain intensity and somatisation, but low depression levels, were key factors associated with the predictive value of CP to indicate symptoms of discogenic origin on positive discography. Specifically, it was found that positive discography and a diagnosis of

discogenic pain were highly likely when centralisation was observed using a McKenzie assessment (MII, 2005) in both non-distressed and not severely disabled (Sp 89%), and distressed and severely disabled (Sp 100%), CLBP patients.

It was concluded that CP observed in non-distressed and not severely disabled CLBP patients examined using a McKenzie assessment (Sp of 89%) is sufficiently diagnostic of discogenic pain. In distressed and severely disabled cases (Sp of 100%), the presence of CP is merely suggestive. It was noted that inability to tolerate the repeated movement testing of the McKenzie assessment was increased in cases of high distress.

In this study, only the 'Flow and Timing' domain of the QUADAS-2 indicated any risk of bias, which was high. Although all patients received the same reference standard, not all patients received the reference standard or were included in the analysis, and the interval between the index tests and the reference standard was not stated.

The second publication by Laslett, Aprill, et al. (2006) investigated the diagnostic accuracy of clinical variables in relation to provocation discography as a reference standard for pain-sensitive intervertebral discs in CLBP. One physiotherapist competent in the McKenzie assessment method conducted a clinical examination synonymous with the McKenzie Institute Lumbar Spine Assessment (MII, 2005), plus additional patho-anatomical provocation tests (see Table 2.2). A radiologist conducted provocative discography as the reference standard to identify the presence or absence of discogenic pain.

Results indicated that Sn was low (58%) for the individual variables of CLBP history, moderate observed loss of lumbar extension and reported 'vulnerability' in the neutral zone, with minor variations according to level of distress and disability.

However, in those with severe disability and distress, Sp was high when extension loss was identified and there was a history of persistent pain, which indicated their value to predict positive discography and thus the presence of discogenic pain. Further, when combined in the presence of CP, these variables showed the highest Sn (78%) and Sp (59%), with +ve LR of 1.9 (1.1–3.2) and −ve LR of 0.37 (0.21–0.65). It was concluded that, in CLBP patients with a positive discogram, signs of discogenic pain include the presence of CP combined with history of persistent LBP, loss of lumbar extension or reported 'vulnerability'.

In this study, the 'Index Test' and 'Flow and Timing' domains of the QUADAS-2 indicated risk of bias. Although an appropriate interval between the index tests and the reference standard was stated, no threshold was pre-specified for the index tests, and not all patients received the reference standard or were included in the analysis.

2.4 Discussion

The identification of the most diagnostically accurate examination processes or tests for physiotherapy CLBP assessment was a fundamental requirement to guide the construction of a comprehensive physiotherapy CLBP examination algorithm that was evidence based for use in some studies included in this research (see Chapters 6 & 7). However, this review identified minimal research to support the concurrent diagnostic accuracy and predictive value of existing physiotherapy examination processes or classification systems used in contemporary clinical practice when applied specifically to CLBP patients. Only five eligible publications reported the diagnostic classification properties of some individual examination tests or combined test clusters, and none evaluated the concurrent diagnostic accuracy of any composite physiotherapy classification system in CLBP.

The reviewed articles reported diagnostic or predictive accuracy of individual and clustered examination tests for discogenic pain (Laslett, Aprill, et al., 2006; Laslett, Öberg, et al., 2005), NR impingement of discogenic origin (Iversen et al., 2013) or FJS (Laslett, McDonald, et al., 2006; Laslett et al., 2004). Although past reviews have reported diagnostic classification and clinical prediction, they were not exclusive to CLBP. This is the first known review to investigate the diagnostic accuracy of physiotherapy examination and classification processes specific to CLBP. This review provided information to guide the development of a physiotherapy examination algorithm for use in this CLBP research. It highlighted that diagnostic accuracy of any existing classification system or process has yet to be established for CLBP and, therefore, clinical application remains unsupported.

All five studies appraised using the QUADAS-2 were highly applicable for the purpose of this literature review. The patients in each of the five studies matched the review's target population of CLBP, participant demography, study setting, index tests conducted and data analysis.

2.4.1 Predictors of discogenic pain

Two studies evaluated examination processes to identify a pain-sensitive disc (Laslett, Aprill, et al., 2006; Laslett, Öberg, et al., 2005). The earliest of these studies concluded that the presence of CP assessed using a McKenzie assessment has diagnostic accuracy to identify CLBP of discogenic origin (Laslett, Öberg, et al., 2005). The later study determined that chronic discogenic pain could be identified in the presence of CP and reduced extension or 'vulnerability', but that CP was the primary clinical indicator (Laslett, Aprill, et al., 2006).

Compared with the FJS-related studies (Laslett, McDonald, et al., 2006; Laslett et al., 2004), the risk of bias in discogenic studies that used provocative

discography was lower. However, one study did not report if index test thresholds were pre-specified (Laslett, Aprill, et al., 2006). If test thresholds are enhanced or reduced compared with levels generally accepted for clinical application, potentially over- or under-inflated test accuracy may be reported, and this was unable to be ruled out. Despite potential bias identified in the domain of 'flow and timing', overall low risk of bias in the study by Laslett, Öberg, et al. (2005), suggests greater clinical applicability of the test clusters reported. It is therefore recommended that a McKenzie assessment be included as part of a physiotherapy CLBP examination to predict the presence of CP associated with symptoms of chronic discogenic origin.

An underpinning feature of the McKenzie assessment is repeated movement testing, which is integral to the assessment of CP. Such tests are considered highly useful to assess symptom response during examination, and thus guide subsequent patient management (McKenzie & May, 2003). Accordingly, the diagnostic value of CP identified by this systematic review supports the inclusion of the McKenzie assessment (MII, 2005) in a proposed physiotherapy CLBP examination algorithm. However, it was noted that inability to tolerate the repeated movement testing of the McKenzie assessment was increased in cases of high 'distress' (Laslett, Öberg, et al., 2005). Therefore, this should be a consideration for practitioners when using this examination process for CLBP patients.

2.4.2 Predictors of chronic radiculopathy

The study by Iverson (2013) indicated that the combination of SLR, femoral nerve stretch, lower limb myotome and dermatome tests, and tendon reflex tests showed improved accuracy, compared with their individual application. Individually, their diagnostic accuracy was insufficient for clinical use. This study demonstrated one of the lowest risks of bias across the studies reviewed; therefore, the results

reported in this review are clinically applicable. Hence, these tests should be applied only as a cluster, with examination response considered with caution and reasoned within the context of the complete patient presentation.

2.4.3 Predictors of facet joint syndrome

The earlier of the two studies that evaluated examination processes to identify FJS (Laslett et al., 2004) refuted the accuracy of Revel's criteria (Revel et al., 1998). The later study by Laslett, McDonald, et al. (2006) determined that the presence of CP combined with negative results for one or more of age greater than 50 years, symptoms best walking or sitting, paraspinal pain and ER was appropriate to rule out the presence of FJS in CLBP patients, but their ability to detect the presence of the condition did not have sufficient accuracy.

However, in both studies, the participant recruitment process was unclear. Selection bias might have resulted, and diagnostic accuracy or predictive values of the index tests might have been inflated if participants with less 'complex' presentations were specifically excluded from these studies (Laslett, McDonald, et al., 2006; Laslett et al., 2004). Conversely, the reverse might have occurred if participants who were easier to classify or diagnose were selected. Neither study reported if index test thresholds were pre-specified. Although the index tests had dichotomous outcomes, the threshold between a positive or negative response was not explicitly stated; as previously described, this may lead to over- or under-inflated test accuracy, which might have been present in these studies.

Blinding was appropriately conducted in both studies (Laslett, McDonald, et al., 2006; Laslett et al., 2004). Additionally, the facet joint injection reference standard was appropriate, given that no diagnostic 'gold standard' for FJS currently

exists. Facet joint injection, performed by a suitably trained injectionist, is considered an acceptable reference standard (Bogduk, 2004).

Both studies (Laslett, McDonald, et al., 2006; Laslett et al., 2004) failed to include all participants in their analysis. Participants lost through exclusion might have differed from the included sample and resulted in bias of the study results. However, the reasons for exclusion stated were appropriate in both studies. Additionally, verification bias was present in one of the studies that did not apply the reference standard to all participants (Laslett, McDonald, et al., 2006). Nevertheless, applying an invasive reference standard to every participant regardless of whether it is clinically indicated or not is unethical. Consequently, often the exclusion of some participants in such clinical studies may be unavoidable.

Diagnostic facet joint injection has a limited duration of pain relief following the procedure. Therefore, there is a maximum time after the reference standard procedure in which to assess a positive or negative injection response. In one of the studies, it was unclear if the interval between index test and reference standard, or assessment of outcome of the reference standard injection, was appropriate (Laslett et al., 2004). Because no interval was explicitly stated, it cannot be assumed that the reference standard outcome was correct, and overall study outcomes may be erroneous.

A limitation of this review was that the studies were appraised by only one reviewer which could have constituted elements of study selection bias. Further, the strict selection criteria, although formulated to provide a narrow focus of knowledge synthesis, may have limited the greater scope of articles identified.

Notwithstanding, this review highlights evidence that Revel's criteria are invalid, but clinicians may rule out FJS in the presence of CP, when pain is not

persistent, the patient's age is not over 50 years, symptoms are not best walking or sitting, onset pain was not paraspinal, ER is negative, loss of lumbar extension is less than moderate and 'vulnerability' is absent.

2.5 Conclusion

This review highlighted sparse evidence for the diagnostic capabilities of some common physiotherapy examination tests and classification systems when applied to patients with CLBP. Overall, a moderate risk of bias was evident across the studies reviewed. In view of this, the findings suggest cautious support for the utility of the selected CPRs to predict discogenic pain and chronic radiculopathy, and to rule out FJS, when applied to CLBP patients. Accordingly, in the absence of alternative clinical tests that demonstrate high diagnostic accuracy, it may be concluded that repeated movement tests using a standard McKenzie Institute Lumbar Spine Assessment (MII, 2005), combined with clinical test clusters for discogenic pain (Laslett, Aprill, et al., 2006; Laslett, Öberg, et al., 2005), chronic radiculopathy (Iversen et al., 2013) and FJS (Laslett, McDonald, et al., 2006; Laslett et al., 2004), have sufficient diagnostic value for inclusion in a physiotherapy CLBP examination algorithm for this research project.

Key points:

- The validity of clinical examination clusters to predict the presence or absence of discogenic pain, radiculopathy and FJS has been reported in only five studies with CLBP populations.
- The studies showed moderate risk of bias, but no other studies have identified examination components with greater diagnostic accuracy.

- There is evidence to support the clinical application of the clustered tests reported by these studies in this population, and for their inclusion as components of a comprehensive physiotherapy CLBP examination algorithm.

- The application of the clinical examination clusters reported in the reviewed studies is recommended, with results considered with caution and in the context of the whole patient presentation.

2.6 How This Chapter Informed Subsequent Stages of the Research

In terms of diagnostic component validity, the results of this systematic review guided the development of a comprehensive CLBP physiotherapy examination algorithm to be included in the methodology of the research conducted (see Chapters 6 & 7). It was important that the algorithm included examination components with the highest diagnostic value. This was achieved by summary appraisal of current evidence, and algorithm inclusion deduced from the findings of the review. A summary table of CLBP clinical examination clusters deduced from the findings of this review, and considered appropriate for inclusion as part of the algorithm, is detailed below (see Table 2.3). Full details, and components of the final algorithm and its development, are provided in Chapter 4.

Table 2.3

Clinical Examination Clusters for Chronic Facet Joint Syndrome, Discogenic Pain and Radiculopathy Suitable for Inclusion in a

Physiotherapy Chronic Low Back Pain Examination Algorithm

Predictors of facet joint syndrome (Laslett, McDonald, et al., 2006; Laslett et al., 2004)	Predictors of discogenic pain (Laslett, Aprill, et al., 2006; Laslett, Öberg, et al., 2005)	Predictors of Radiculopathy of Discogenic Origin (Iversen et al., 2013)
Absence of centralisation phenomenon on McKenzie examination (MII, 2005).	Presence of centralisation phenomenon on McKenzie examination (MII, 2005).	Presence of centralisation phenomenon on McKenzie examination (MII, 2005).
Age over 50 years.		+ve SLR or femoral nerve stretch.
Localised unilateral paraspinal LBP never referred below the knee.		Lower limb myotome and dermatome deficit.
Pain best when walking or sitting.		Tendon reflex deficit.
MSPQ $\geq$13.		
Positive extension rotation test.		

Note. MSPQ = Modified Somatic Pain Perceptions Questionnaire (Main, 1983); SLR = straight leg raise test.

Chapter 3: Reliability of Physiotherapy Classification in Chronic Low Back Pain

This chapter in combination with Chapters 2 and 4, addresses a component of the first research question (see Chapter 1): 'What physiotherapy lumbar spine examination processes or classification systems have demonstrated sufficient validity, diagnostic accuracy and reliability in CLBP for inclusion in a comprehensive physiotherapy CLBP examination algorithm?' Specifically, this chapter reviews studies that have applied physiotherapy LBP classification systems, incorporating physical examination methods to report inter-rater reliability exclusively in CLBP patients.

This chapter reports a published study entitled 'Inter-Rater Reliability of Classification Systems in Chronic Low Back Pain Populations' (Flavell, Gordon, Marshman, & Watt, 2014; see Appendix 3).

3.1 Introduction

Classification systems for LBP have been described as structured clinical assessment pathways that identify subgroups of patients (Heiss et al., 2004). Physical therapists use LBP classification systems, which include, but are not limited to, evaluation of baseline symptom behaviour, examination of movements and posture, neurological and neuro-dynamic testing, and assessment of spinal stability (Charlin et al., 2012).

Several detailed classifications have been developed, with the belief that subgrouping people with LBP is important for both clinical and research purposes (McCarthy, Rushton, Billis, Arnall, & Oldham, 2006; Saragiotto et al., 2017; Wilde, Ford, & McMeeken, 2007). They include, but are not exclusive to, systems such as mechanical diagnostic therapy (McKenzie, 1981), movement system impairment

classification (MSI; Sahrmann et al., 2003), motor control impairment classification (OCS; O'Sullivan, 2005), classifications based on the distribution of symptoms (Hall, McIntosh, & Boyle, 2009), patho-anatomical origin of symptoms (Ford et al., 2016; Schwarzer et al., 1994), tools that classify LBP patients according to risk of chronicity (Hill et al., 2008), and predictors of outcome (Fritz, Beneciuk, George, Hill, & Hay, 2011; Kongsted & Leboeuf-Yde, 2010).

Valid and reliable patient classification improves clinical outcomes and research methodology, and constitutes best clinical practice by informing targeted interventions, which lead to improved patient outcomes. Predominantly, classifications used by physical therapists incorporate clinical examination techniques. Therapists implement many examination techniques, including manual therapy and movement-based assessments, with responses used for classification and to predict treatment outcome.

Classification is also important as a process of subgrouping to homogenise research participants. This improves methodological rigour and ultimately research outcomes (McCarthy et al., 2006). LBP research participants often have been grouped according to the duration of their symptoms. Three duration-based groups exist: acute, subacute and CLBP. Further subgrouping within the three duration groups, based on movement patterns and patho-anatomical origin, may serve to homogenise LBP populations further.

LBP classification systems usually follow a detailed algorithm and guide treatment decisions via a clinical reasoning process, thereby reducing extraneous information gathering. Effective clinical reasoning is a key aspect of assessment; however, it is one with which undergraduate, newly graduated and inexperienced health professionals struggle (Charlin et al., 2012). Hence, a defined yet flexible

classification system using effective clinical reasoning pathways would facilitate and provide clarity for practitioners.

Reliable classification systems, when applied to homogeneous subgroups, support practitioners to achieve best physiotherapy practice with more specific assessment, and therefore focused interventions and improved positive outcomes for patients. This is particularly important for patients with CLBP, whose symptoms are of longer duration and whose clinical presentation is often complicated by psychosocial factors (McCarthy, Arnall, Strimpakos, Freemont, & Oldham, 2004).

The prevention of chronicity in LBP has been a long-established, yet difficult, goal to attain for health professionals. Physical therapists regularly encounter patients whose LBP has already persisted to a chronic stage of greater than 3 months' duration, and who have received little or no previous intervention by health professionals. Hence, it is proposed that the reliability of existing classification systems in populations with CLBP should be established before recommendations for their use can be made.

A systematic review was conducted to achieve two objectives. (1) To identify and appraise the current literature about the reliability of LBP classification systems when applied in homogeneous populations of CLBP patients. (2) To identify the most reliable classification system in a CLBP population.

3.2 Method

This systematic review was registered with the international prospective register of systematic reviews (CRD42013003655).

3.2.1 Search strategy

The following databases were accessed: Medline via OvidSP (1946 to September 2013); CINAHL (no date restriction); PEDro (no date restriction); the

Cochrane library (no date restriction); Informit (1970 to September 2013); and Scopus (no date restriction). An initial keyword, title and abstract search using the following search terms (truncated as required) was conducted: LBP; diagnosis; risk; classification; algorithm; develop; screening; and reliability. A further search strategy was incorporated with the following key terms: mechanical diagnosis and treatment; treatment-based classification; patho-anatomical classification; movement system impairment classification; O'Sullivan classification system; and motor control impairment. The Boolean operators 'and' and 'or' were applied between each search term (see Table 3.1). The chief reviewer (CAF) conducted the database search on 9 September 2013.

3.2.2 Study selection

Potentially suitable articles were identified from the title and imported into EndNote (version 16). The chief reviewer (CAF) and second reviewer (SG) reviewed abstracts of identified articles. The reference lists of the identified abstracts were scanned for further suitable articles. Upon agreement, full texts were sourced for inclusion by both reviewers. Disagreement between reviewers was resolved via consensus. Both reviewers appraised full text articles for inclusion according to the following criteria.

3.2.3 Eligibility criteria

3.2.3.1 Types of studies

Inter-rater reliability studies of LBP classification systems that incorporated physical examination methods.

Table 3.1

Search Strategy

1	Low back pain*AND Classification*	AND	Diagno*	AND	Reliability*				
2	Low back pain*AND Classification*	AND	Algorithm*	AND	Reliability*				
3	Low back pain*AND Classification*	AND	Reliability*	AND	Develop*				
4	Low back pain*AND Algorithm*	AND	Diagno*	AND	Reliability*				
5	Low back pain*AND Risk*	AND	Screening*	AND					
6	Low back pain*AND Risk*	AND	Reliability*						
7	Low back pain*AND Screening*	AND	Reliability*						
8	Low back pain*AND Screening*	AND	Reliability*		Algorithm*				
9	Mechanical diagnosis and treatment*Or Treatment-based classification *	Or	Patho-anatomical classification *	Or	Movement system impairment classification*	Or	O'Sullivan classification svstem*	Or	Motor control impairment
10	Low back pain*AND Mechanical diagnosis and treatment*	Or	Treatment-based classification *	Or	Patho-anatomical classification *	Or	Movement system impairment classification*	Or	O'Sullivan classification system*

3.2.3.2 Participants

All study participants were required to be adult (≥18 years old), exclusively with LBP defined as pain between the level of the 12th thoracic vertebra and the buttock crease, with or without associated lower limb symptoms (Jone et al., 2003), persisting for more than 12 weeks.

3.2.3.3 Type of outcome measure or intervention

Studies that reported detailed reliability statistics for one or more LBP classification system were eligible for review.

3.2.4 Exclusion criteria

Systematic reviews, unpublished non-peer-reviewed publications, opinion pieces, discussion papers and studies not published in English were excluded. Exclusion applied to studies that included any participants with symptoms of less than 12 weeks' duration (acute and subacute LBP). Articles were also excluded for review if participants were pre- or post-partum, had been diagnosed with inflammatory disease, malignancy or pain of visceral origin, or were postoperative spinal surgery patients. Additional exclusions were any system based solely on clinical observation, self-reported questionnaires or other non-physical examination methods.

3.2.5 Data extraction process and review of methodological quality

Both reviewers extracted data from the eligible studies and appraised each using the quality appraisal of reliability studies (QAREL) data extraction form and checklist (Lucas, Macaskill, Irwig, & Bogduk, 2010). The QAREL is a study appraisal tool designed specifically for reliability studies to evaluate risk of bias for both internal and external validity, and statistical analysis (Lucas et al., 2010).

The topics included in this appraisal tool are participants, raters, blinding examination order, application and timing of tests, risk of bias and use of appropriate statistics. Articles were scored as 'yes', 'no' or 'unclear'. Some sections could be scored 'not applicable'. 'Yes' indicated good quality, and 'no' poor quality. The QAREL has been used in previous systematic reviews of reliability studies (Adhia, Bussey, Ribeiro, Tumilty, & Milosavljevic, 2013; Carlsson & Rasmussen-Barr, 2013).

Both reviewers (CAF & SG) scored the final full text articles independently. Reviewers discussed and set the acceptable benchmarks for rating blinding and stability of variable sections on the QAREL checklist. Following independent review, any disagreement was resolved by consensus.

3.2.6 Synthesis and analysis of results

QAREL outcomes were summarised to allow comparison of study quality. In agreement with previous studies, seven 'yes' QAREL checklist items indicated a moderate risk of bias. Consequently, less than seven 'yes' items indicated a high risk of bias, and eight or more a low risk of bias, hence indicating the study to be of good quality (Lucas et al., 2010; Simopoulos et al., 2012). Internal and external validity were scored separately, and 'yes' scores were calculated as a percentage of possible scores in that section. For each section (internal and external validity), 67% or greater defined the benchmark level for high quality, 50% or greater moderate, and less than 50% poor quality in the studies (Simopoulos et al., 2012; van der Wurff, Hagmeijer, & Meyne, 2000).

3.3 Results

3.3.1 Study selection

Using the search syntax previously described (see Table 3.1), the reviewers identified 2384 research articles from the electronic databases. Screening and review were conducted according to the standardised Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (Moher et al., 2009; see Figure 3.1). Subsequent to title screening, to inclusions from reference list scanning and to removal of duplicate articles, 84 studies were identified for abstract screening. Both reviewers (CAF & SG) assessed the abstracts independently. Reviewer disagreement on seven abstracts was resolved by consensus agreement. Twenty-two abstracts were identified as suitable for full text review. Both reviewers independently scanned the full text versions and screened for eligibility.

Following full text review, 19 articles were rejected. Reasons for rejection included studies that did not evaluate reliability ($N = 6$; Ferguson, Gallagher, & Marras, 2003; Fritz et al., 2011; Main, Sowden, Hill, Watson, & Hay, 2012; Marras et al., 1999; McCarthy, Roberts, Gittins, & Oldham, 2012; Sahrmann et al., 2003), did not evaluate CLBP exclusively ($N = 10$; Bertilson, Bring, Sjöblom, Sundell, & Strender, 2006; Clare, Adams, & Maher, 2004; Henry, Van Dillen, Trombley, Dee, & Bunn, 2013; Hill, Vohora, Dunn, Main, & Hay, 2010; Petersen et al., 2004; Razmjou, Kramer, & Yamada, 2000; Riddle & Rothstein, 1993; Vibe Fersum, O'Sullivan, Kvåle, & Skouen, 2009; Widerström, Olofsson, Arvidsson, Harms-Ringdahl, & Larsson, 2012; Wilson, Hall, McIntosh, & Melles, 1999), did not evaluate a clinical examination process for classification ($N = 1$; Ferguson et al., 2003) or were classified as a prediction of risk for chronicity ($N = 2$; Fritz et al., 2011; Hill et al., 2008).
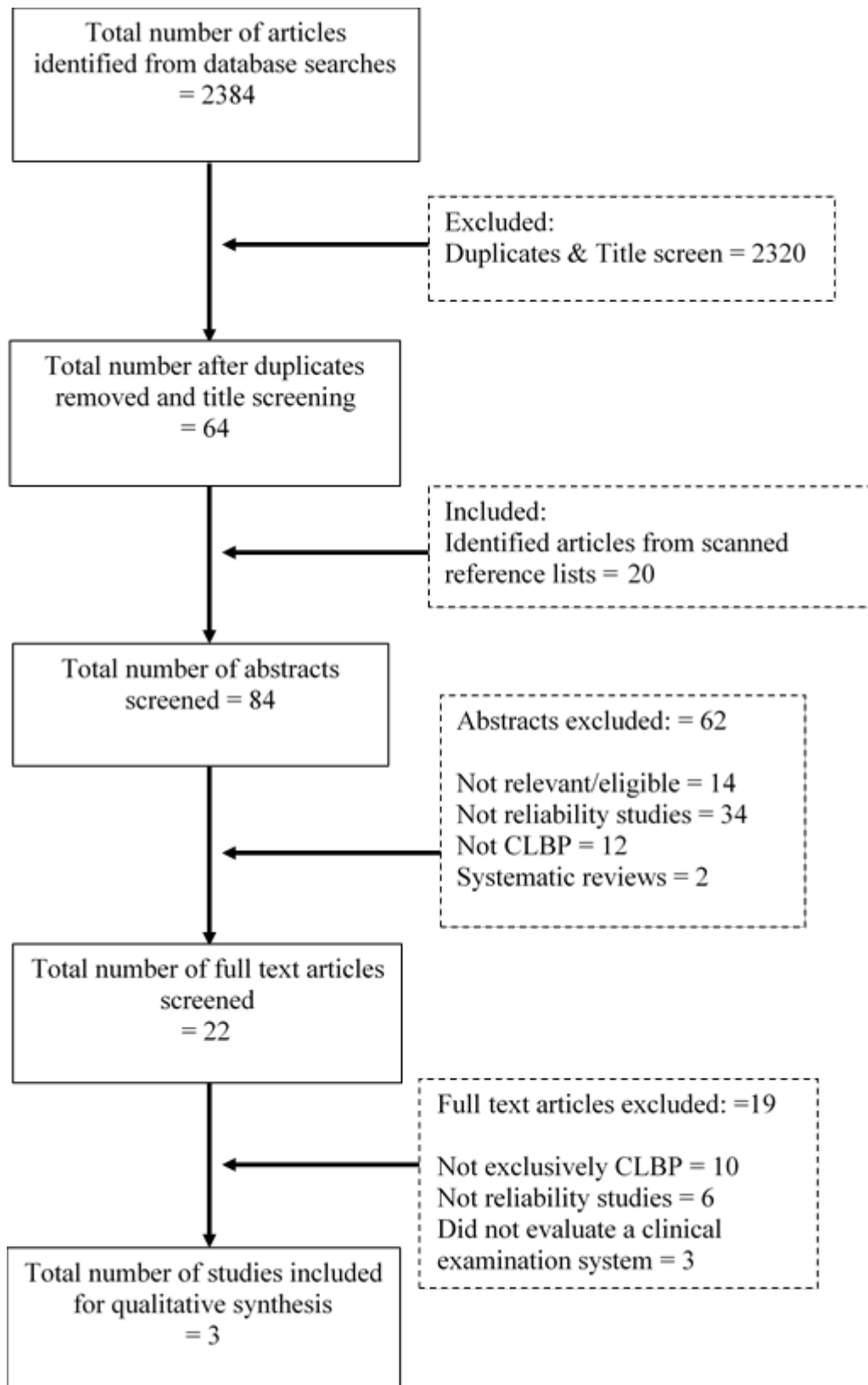
Figure 3.1. Flow chart of study selection process.

Several studies reported on reliability of heterogeneous LBP populations, but the number of participants in each symptom duration subgroup was not specified, making it impossible to evaluate the results specific to CLBP ($N = 8$; Bertilson et al., 2006; Henry, Fritz, Trombley, & Bunn, 2012; Kilpikoski et al., 2002; Petersen et al., 2004; Razmjou et al., 2000; Riddle & Rothstein, 1993; Widerström et al., 2012; Wilson et al., 1999). In addition, the definition of CLBP was sometimes ambiguous ($N = 1$; Razmjou et al., 2000). One further study specified the number of participants according to symptom duration, but was rejected because it evaluated a questionnaire-based tool, which did not include a physical examination component (Hill, Vohora, et al., 2010).

The reviewers initially disagreed on the eligibility of three of the 19 excluded studies (Kilpikoski et al., 2002; Marras et al., 1999; Razmjou et al., 2000). The authors of the three contentious articles were contacted to clarify details of the classification process (Marras et al., 1999) or whether significant findings from the CLBP participants in the study population could be reported (Kilpikoski et al., 2002; Razmjou et al., 2000). Responses from two of the three author groups were received. Subsequently, and in consideration of all available information, consensus via discussion was reached by the reviewers. Three full text articles of studies conducted exclusively on CLBP populations were accepted for final review. A description of the studies is provided in Table 3.2. Two classification systems were evaluated in the studies accepted for review: inter-rater reliability of the MSI (Harris-Hayes & Van Dillen, 2009; Trudelle-Jackson, Sarvaiya-Shah, & Wang, 2008) and the OCS (Dankaerts, O'Sullivan, Straker, Burnett, & Skouen, 2006). The study by Dankaerts et al. (2006) consisted of two separate reliability studies, which were evaluated as 'study 1' and 'study 2'.

3.3.2 Risk of bias within studies

The QAREL checklist of 11 items was applied to each of the four studies. Disagreement between the two reviewers occurred on five single items (11%). Disagreement was resolved by consensus discussion. Results of the QAREL evaluation for risk of bias ranged between six (low) and eight (high; see Table 3.3). Study 1 by Dankaerts et al. (2006) assessed the reliability of 'expert' raters ($N = 2$) to classify CLBP using the OCS. Study 2 by Dankaerts et al. (2006) assessed the reliability of raters who were 'moderately familiar' and 'very familiar' with the OCS ($N = 17$). Study 1 showed a high risk of bias but study 2 a low risk of bias. Both studies that used the MSI showed a high risk of bias (Harris-Hayes & Van Dillen, 2009; Trudelle-Jackson et al., 2008).

Internal validity, external validity and statistical methods were evaluated using the QAREL. The external validity of all studies was considered high. However, internal validity varied between classifications, with the MSI studies by Harris-Hayes and Van Dillen (2009) and Trudelle- Jackson et al. (2008) showing an internal validity of 60%, compared with study 1 and study 2 by Dankaerts et al. (2006) of the OCS, which showed internal validity values of 80% and 100%, respectively. Statistical analysis rated highly (100%) for both classification systems, except in study 1 by Dankaerts et al. (2006).

3.3.3 Summary of results

All studies examined inter-rater reliability, and kappa values ranged from .32 to .96 (see Table 3.2). Percentage of agreement was reported in all studies and ranged between 44% and 97%. The results of Dankaerts et al. (2006) study 1 conducted with two expert raters showed almost perfect agreement ($k = .96$, 97% agreement) but with a high risk of bias.

Table 3.2

Summary Description and Results of Included Articles

Study	Objective	Classification system	Population	Raters	Method	Statistical analysis	Results
Dankaerts et al. (2006) (study 1)	Determine inter-rater reliability of expert raters for classification of participants with NS-CLBP with OCS.	Motor control impairment classification	$N = 35.18$ women. All NS-CLBP. Mean duration 5.6 yr. Age 37 (12.73) yr. ODI (%) 37 (11).	Two musculoskeletal physical therapists. One the developer of the system and 18 yr experience in LBP. One extensive training by the developer and 12 yr experience in LBP.	Raters were blinded. Re-examination 24 hr to 1 week.	Five diagnostic outcome variables (categorical) Kappa (*k*) & percentage agreement.	Agreemen*t* (%) = 97 *k* = .96

Note. NS-CLBP = non-specific chronic low back pain; OCS = motor control impairment classification; ODI = Oswestry Disability Index (Fairbank & Pynsent, 2000); GP = general practitioner.

Table 3.2

Summary Description and Results of Included Articles (Continued)

Study	Objective	Classification system	Population	Raters	Method	Statistical analysis	Results
Dankaerts et al. (2006) (study 2)	Determine inter-rater reliability of clinicians for classification of participants with NS-CLBP with OCS.	Motor control impairment classification	$N = 25$. Summary not given.	$N = 8$ raters. 1 GP, 1 clinical neurologist, 3 musculoskeletal physical therapists, 2 physical therapists. Moderately familiar with OCS. Training by clinical workshop with developer and instruction package $N = 5$ raters. 4 musculoskeletal physical therapists, 1 sports physical therapist. Very familiar with classification system. Postgraduate training by developer.	Video and case reports evaluated. Raters were blinded. Initially rated on case report only. Video and case reports evaluated. Raters were blinded. Initially rated on case report only. Followed by combined case report and video examination.	Five diagnostic outcome variables (categorical) Kappa (k) & percentage agreement between expert raters (gold standard) and other raters. Additionally, agreement with case report only or combined information.	1. All clinicians: a. Case report only Agreemen$t$ (%) = 48 k = .32 b. Case report & video Agreemen$t$ (%) = 70 k = .61 2. Between clinician types: Moderately familiar a. Case report only Agreemen$t$ (%) = 44 k = .28 b. Case report & video Agreemen$t$ (%) = 65 k = .55 Very familiar a. Case report only. Agreemen$t$ (%) = 54 k = .40 b. Case report & video Agreemen$t$ (%) = 78 k = .71

Note. NS-CLBP = non-specific chronic low back pain; OCS = motor control impairment classification; ODI = Oswestry Disability Index (Fairbank & Pynsent, 2000); GP = general practitioner.

Table 3.2

Summary Description and Results of Included Articles (Continued)

Study	Objective	Classification system	Population	Raters	Method	Statistical analysis	Results
Trudelle-Jackson et al. (2008)	Determine inter-rater reliability for classification of participants with CLBP.	Movement system impairment classification	$N = 24$. 16 women. All CLBP (>12 weeks). Mean duration 288 weeks. Age 43.8 (13.5) yr ODI (%) 37.4 (17.8).	Two physical therapists. Experience varied. Both trained in system use via courses. One rater trained by system developer. Raters practised together on student subjects >8 hr over 2 weeks prior to study.	Raters were blinded. 25 test items. Examinations conducted sequentially. No rest period and same day.	Five diagnostic outcome variables (categorical) Kappa (k) & percentage agreement.	Agreement (%) = 75 $k = .61$ ($p < .001$) 95% CI [0.33, 0.89]
Harris-Hayes and Van Dillen (2009)	Determine inter-rater reliability for classification of participants with LBP.	Movement system impairment classification	$N = 30$. 21 women. No duration of symptoms reported. Stated as CLBP. Mean age 31.1 (12.9) yr. ODI (%) 13.6 (7.5).	Two physical therapists, both with >10 yr musculoskeletal experience. One rater was the developer of the system; the other had continuing education and 7 yr experience using the system. Training for the study was conducted by 2nd rater. Study of operations manual and practice with symptomatic and asymptomatic subjects.	Raters were blinded. Examination on same day with 15 minute break between.	Five diagnostic outcome variables (categorical) Kappa (k) & percentage agreement.	Agreement (%) = 83 $k = .75$ ($p < .001$) 95% CI [0.51- 0.99] $Z = 6.17$

Note. NS-CLBP = non-specific chronic low back pain; OCS = motor control impairment classification; ODI = Oswestry Disability Index (Fairbank & Pynsent, 2000); GP = general practitioner.

Table 3.3

Summary of Risk of Bias Evaluation Using Quality Appraisal of reliability Studies Checklist

Diagnostic/Classification approach		Dankaerts et al. (2006)		Trudelle-Jackson et al. (2008)	Harris-Hayes and Van Dillen (2009)
		Study 1	Study 2		
Q 3-9: Internal Validity items	Blinding: inter-rater	Y	Y	Y	Y
	intra-rater	NA	NA	NA	NA
	from reference standard	NA	Y	NA	NA
	from other clinical information	Y	Y	Y	Y
	from other clues	U	Y	U	U
	Variation of examination order	Y	NA	Y	Y
	Suitable time interval between tests/ examinations	Y	Y	N	U
	Total	4/5=80%	5/5=100%	3/5=60%	3/5=60%
Q1,2 & 10: External Validity items	Suitable participant sample	Y	Y	Y	Y
	Suitable raters	Y	Y	Y	Y
	Appropriate test/examination conducted	U	U	U	U
	Total	2/3=66.66 %	2/3=66.66%	3/3=100%	2/3=66.66%
Q11: Statistics	Total	0/1=0%	1/1=100%	1/1=100%	1/1
Overall Total number 'Yes'		6	8	6	1/1=100%
Risk of Bias		High	Low	High	High

Note. Y = yes; N = no; U = unclear; NA = not applicable. *Sourc*e*:* Lucas et al. (2010).

In study 2 of the same article, when 'moderately familiar' and 'very familiar' raters reviewed subjective case reports plus video examination, the results showed moderate ($k = .55$, 65% agreement) and substantial ($k = .71$, 78% agreement) agreement, respectively. When the same raters reviewed subjective case reports only, fair agreement was reported for both 'moderately familiar' ($k = .28$, 44%) and 'very familiar' ($k = .4$, 54% agreement) raters. Agreement for the MSI was substantial in both studies ($k = .75$, 83% agreement; Harris-Hayes & Van Dillen, 2009 and $k = .61$, 75% agreement; Trudelle-Jackson et al., 2008).

3.4 Discussion

This review identified that only the reliability of the OCS and MSI when applied to homogeneous populations with CLBP has been reported. A previous review about reliability of LBP classification systems focused on non-specific LBP did not evaluate study bias, and the research was not exclusive to any particular duration-based subgroup of non-specific LBP (Karayannis, Jull, & Hodges, 2012).

The two independent reviewers conducted a pre-appraisal discussion to reach agreement on the key aspects of the QAREL checklist as recommended by Lucas et al. (2010); 80% of disagreement was on item eight of the 11-point list, 'Was the order of examination varied?' This was clarified during the pre-appraisal discussion, but reporting consensus between reviewers remained challenging. The authors are not aware of any research related to the reliability of the QAREL appraisal tool or any of its individual items despite its use in previous studies. Hence, the authors of this review support previous suggestions by Lucas et al. (2010) that further investigation be conducted to evaluate the reliability of the QAREL.

With the exception of Dankaerts et al. (2006) study 2, all articles showed a high overall risk of bias. High external validity was reported for both the OCS and

MSI. However, internal validity was lower for the MSI (Harris-Hayes & Van Dillen, 2009; Trudelle-Jackson et al., 2008) than the OCS. This was related particularly to lack of blinding from extraneous cues apparent when consecutive face-to-face examinations by two raters were conducted, compared with examiners' classifications from pre-recorded videos (Dankaerts et al., 2006, study 1; Harris-Hayes & Van Dillen, 2009; Trudelle-Jackson et al., 2008). For example, in the MSI studies, examinations were conducted only 15 min apart (Trudelle-Jackson et al., 2008) or with an unspecified time interval (Harris-Hayes & Van Dillen, 2009), and participants could potentially memorise previous symptom responses, resulting in recall bias. Therefore, the authors of this review believe that face-to-face re-examination conducted in quick succession might have led to recall bias and low internal validity in the studies by Trudelle-Jackson et al. (2008), Harris-Hayes and Van Dillen, (2009), and Dankaerts et al. (2006) study 1, compared with the study using only pre-recorded videos (Dankaerts et al., 2006, study 2).

The results of this review indicated that where raters were either 'experts' or 'very familiar' with the classification system, inter-rater reliability was substantial or almost perfect. These findings for CLBP are consistent with those of Fairbank et al. (2011), who concluded that rater training contributed significantly to reported reliability in studies of LBP classification systems. This has important implications for clinical practice. It highlights the need for continued professional development and specific training in the utility of these systems. Thereby, clinicians will maximise reliability when applying classification systems and subsequently target treatments for CLBP patients appropriately.

Variations in examination methodologies, re-examination time intervals, and experience and training in the application of the classification systems limited

comparison between studies, and the conclusions from this review. A similar paucity of methodological consistency is evident among reliability studies, which investigated different LBP classifications in cohorts of mixed LBP duration (Clare, Adams, & Maher, 2005; Fritz, Brennan, Clifford, Hunter, & Thackeray, 2006; Heiss et al., 2004; Henry et al., 2012; Riddle & Rothstein, 1993; Widerström et al., 2012; Wilson et al., 1999). This identifies the need for consensus on inter-rater reliability study design and reporting.

Although the importance of preventing progression to chronicity in LBP has been discussed previously (Fritz et al., 2011; Hill et al., 2008; Hill, Dunn, Main, & Hay, 2010; Hill, Vohora, et al., 2010; Widerström et al., 2012), circumstances prevail whereby some LBP patients first access physiotherapy in a chronic state. Therefore, improved methods to classify patients with CLBP are urgently needed. Limitations of this study include exclusion of articles not published in the English language. Nevertheless, this review established that minimal high quality evidence for inter-rater reliability of CLBP classification exists.

3.5 Conclusion

This review indicated that there is a lacuna of research reporting the inter-rater reliability of LBP classification systems when applied to CLBP populations. Currently, the inter-rater reliability of only two systems, the OCS and MSI, has been investigated in this population. The clinical implications of this review are that there is a lack of evidence for the reproducibility of these classifications for clinical use in this population. Hence, research outcomes of randomised controlled trials utilising these classifications systems to study CLBP populations remain unsupported until risk of bias is reduced and reliability is established. Accordingly, to facilitate improved classification, management and outcomes for these patients, there is a

critical need to conduct research on the reliability of existing LBP classification systems in CLBP populations.

Key points:

- The inter-rater reliability of only two systems, the OCS and MSI, has been reported specifically for CLBP.

- There is a lack of evidence to support the clinical application of any existing LBP classification system as a 'standalone' examination process in this population.

- There is a critical need to conduct further research to establish reliable and valid comprehensive CLBP physiotherapy examination processes to address the second and third research questions (see Chapter 1).

3.6 How This Chapter Informed Subsequent Stages of the Research

It was important that the examination process proposed for this research should be as reliable as possible. The results of this systematic review indicated that no existing classification system would provide a reliable, valid or comprehensive algorithm for assessment and classification of CLBP in this research (see Chapters 6 & 7).

Chapter 4: Development of a Physiotherapy Chronic Low Back Pain Examination Algorithm (MK-C)

This chapter addresses the first research question (see Chapter 1): 'What physiotherapy lumbar spine examination processes or classification systems have demonstrated sufficient validity, diagnostic accuracy and reliability in CLBP for inclusion in a comprehensive physiotherapy CLBP examination algorithm?' Specifically, this chapter presents the development of a comprehensive physiotherapy CLBP examination algorithm, which includes information gained from Chapters 2 and 3.

4.1 Introduction

The process to identify valid CLBP examination components most suitable for inclusion in an examination algorithm, and the order of the examination process, followed the stage-based approach previously outlined (see Chapter 1). The systematic reviews of examination and classification validity and reliability (see Chapters 2 & 3) identified moderate validity for some examination clusters when applied in CLBP. From this evidence, a comprehensive physiotherapy CLBP examination algorithm was developed.

4.2 Key Components

Common to all physiotherapy examination is the structured process of an initial history and physical examination, culminating in a final diagnosis. The type of LBP classification system currently used by physiotherapists often varies according to the physiotherapist's training and personal preference (Hill et al., 2008; McKenzie, 1981; O'Sullivan, 2005; Petersen et al., 2003; Sahrmann et al., 2003). Some researchers have proposed classification algorithms to detect subgroups of

LBP (Laslett & van Wijmen, 1999), where treatment application is based on the system of diagnosis and not the therapist's particular system preferences (Petersen et al., 2003). These have encompassed a biopsychosocial approach to LBP examination rather than a single paradigm such as movement, function or motor control. However, no examination process has been developed specifically for CLBP, and existing LBP classification systems or algorithms have not demonstrated validity and reliability when applied in CLBP populations (Flavell et al., 2014).

It is important that a physiotherapy CLBP examination algorithm embrace a biopsychosocial approach to diagnostic classification, because a combination of biological, psychological and social factors is known to influence the development of CLBP (Jull & Moore, 2012; van der Hulst, Vollenbroek-Hutten, & Ijzerman, 2005). The clinical application of a biopsychosocial model was first described by Engel (1981). This approach considers biological, psychological and social factors, and how these aspects of patient presentation interact, helping health professionals understand illness in its complete context (Engel, 1981). This approach is believed to deliver more appropriate patient-centred treatment via a thorough understanding of illness processes (Alonso, 2004).

In developing a comprehensive CLBP physiotherapy examination algorithm, a biopsychosocial framework for examination using physical test components should be combined with an assessment of social and psychological status. This should be appropriate to the context of CLBP, and within the scope of a physiotherapy assessment.

4.2.1 Initial history

The principles of patient history taking during a spinal assessment have been documented and discussed extensively (Hengeveld & Banks, 2014; Petty, 2013;

Refshauge & Gass, 2004). A stepped process that includes completion of a body chart, elucidating behaviour of LBP symptoms, special questioning related to 'red flag' identification, adverse psychosocial influences, history of the LBP, past medical history, and social and family history has been advocated (Petty, 2013). The development process for recording symptom duration, pain intensity and social history taking are discussed in section 4.4 of this chapter.

A body chart, included during history taking, will indicate the location, description, constancy, depth and relationships among current symptoms, and has reported value in clinical diagnosis (von Baeyer, Lin, Seidman, Tsao, & Zeltzer, 2011). Information about the behaviour of symptoms indicates severity, irritability and nature of symptoms, and guides the extent of the physical examination (Hengeveld & Banks, 2014). This process should include but not be limited to questions regarding aggravating and easing factors, 24-hour pattern of symptoms and whether symptoms are improving, worsening or unchanging (Petty, 2013).

A final section of the history taking will include LBP specific questions, to rule out or indicate serious pathologies or conditions with potential to masquerade as mechanical LBP, and which may limit or contraindicate a physical examination. Other questions should explore current and past medications, and the results of any diagnostic investigation or imaging previously conducted (Petty, 2013).

For the purposes of this research, it was important that the physiotherapy CLBP examination algorithm represented a comprehensive process. Therefore, it was considered appropriate that the history-taking component of the algorithm should include each of the elements described herein.

4.2.2 Psychosocial assessment

A variety of self-reported questionnaires are available for use in LBP examination, primarily to evaluate level of disability, function, general well-being and quality of life (Morris, Hee, Stallard, Underwood, & Patel, 2015), and provide an outcome measure to compare baseline scores against treatment progress. Other questionnaires are available to quantify the presence of psychological distress levels and pain somatisation. They contribute to the assessment of adverse psychosocial factors, and indicate when onward referral for psychological assessment is required.

The recording of symptom duration, inclusion of a self-reported pain intensity assessment and a measure of pain somatisation were considered essential physiotherapy CLBP examination inclusions. Pain intensity assessment is important, as high levels of pain have been associated with poor treatment outcome in CLBP (van der Hulst et al., 2005). The 11-point VAS was chosen because it has good reliability and higher sensitivity for reporting pain than do other measures for chronic joint pain (Hawker, Mian, Kendzerska, & French, 2011; Huskisson, 1974). Similarly, recording symptom duration (months) has significance for patient assessment in CLBP because of its association with pain intensity and functional outcome at the 12-month follow-up (Dunn & Croft, 2006).

The MSPQ (Main, 1983; see Appendix 1C), as a measure of pain somatisation and distress, is also relevant and appropriate for use in a CLBP examination algorithm as it was devised and evaluated specifically for CLBP, and an MSPQ score greater than 13 is a component of the clinical test cluster for lumbar FJS (Laslett, McDonald, et al., 2006). In addition, it is unlike most other psychological tests for distress, which are not sensitive in CLBP (Main, 1983; Main, Wood, Hollis, & Spanswick, 1992).

It was also considered vital that self-reported disability measurements were included in the physiotherapy CLBP examination algorithm, as a strong relationship among LBP, fear-avoidance behaviour and disability has been reported (Guclu et al., 2012). For this purpose, the ODI (Fairbank, Couper, Davies & O'Brien, 1980; Fairbank & Pynsent, 2000) and the RM (Roland & Morris, 1983; see Appendix 1A & B) were chosen because they have high sensitivity for reporting disability in CLBP (Leclaire, Blier, Fortin, & Proulx, 1997), as well as validity and good reliability for severe and mild LBP symptoms (Fairbank & Pynsent, 2000).

Certain sociodemographic factors have been associated with CLBP. Specifically, the demographic indicators of age (years), gender and BMI (Cimolin et al., 2011; DePalma, Ketchum, & Saullo, 2012; Shiri, Karppinen, Leino-Arjas, Solovieva, & Viikari-Juntura, 2010) were considered appropriate for inclusion in the physiotherapy CLBP examination algorithm, because variations in these characteristics relate to particular sources of CLBP or are contributing factors (DePalma et al., 2012).

Determining the patient's 'social state' by recording and assessing usual environment (residential and social) is also an important part of the psychosocial examination component. 'Social state', which includes but is not limited to the physical environment, home activities, care requirements and occupation, like the self-reported questionnaires is best recorded after history of the condition has been taken (Campbell & Szmukler, 1993), and thus prior to the physical examination.

4.2.3 Physical examination

The systematic review previously conducted (see Chapter 2) indicated that when the patient history and repeated movement testing were positive for DP with CP as defined by McKenzie and May (2003), and May and Aina (2012), this may

indicate the presence of discogenic CLBP. Further, the systematic review identified that combined clustered clinical tests validated in CLBP included clinical assessment for CP to determine the presence or absence of discogenic pain (Laslett, 2008; Laslett, Aprill, et al., 2005; Laslett, McDonald, et al., 2006; Laslett et al., 2003). On the basis of this premise, it was appropriate to include an examination of quality and range of individual lumbar spine movements followed by repeated movement tests as described by McKenzie and May (2003), and May and Aina (2012), into the CLBP algorithm.

The use of end-range loading through repeated movement testing is exemplified by the McKenzie assessment (MII, 2005) mentioned previously (see Chapter 2.3.4.2). Since its early development, this assessment has been incorporated into examination algorithms that have included a series of patho-anatomical diagnostic tests for FJS, SIJS, spinal instability (SI), spinal stenosis and myofascial presentations (Eirikstoft & Kongsted, 2014; Laslett & van Wijmen, 1999; Petersen et al., 2003). The preeminent factors for the inclusion of the McKenzie assessment in these examination algorithms was founded on its ability to rule out non-mechanical low back symptoms, assist differential diagnosis of mechanical LBP, and identify CP as an indicator of symptomatic discogenic lumbar spine pathology.

The inclusion of elements into the comprehensive CLBP physiotherapy examination algorithm to facilitate differential diagnosis was fundamental to its development. Womersley and May (2006) demonstrated the clinical value of the McKenzie Institute Lumbar Spine Assessment (MII, 2005) to identify patients with or without LBP symptoms related to postural causes, and discussed the relevance of this to differential diagnosis. Notwithstanding, it might be argued that the inclusion of assessment for postural-related symptoms in the CLBP population would be

redundant, given the syndromes low prevalence, younger age demographic and absence of constant presenting symptoms (McKenzie & May, 2003). Conversely, the findings of Womersley and May (2006) supported an opposing view. These authors recognised that its role in differential LBP diagnosis was important to support differentiation of mechanical from non-mechanical presentations, and to make appropriate classification. Thus, the capability of the McKenzie Institute Lumbar Spine Assessment (MII, 2005) to facilitate differential diagnosis provided further indication for its inclusion as a component of the physiotherapy CLBP examination algorithm.

It was necessary that the developed CLBP examination algorithm also include an evaluation of neurological and neuro-dynamic status relevant to LBP. Such an evaluation may include lower limb myotome, dermatome and tendon reflex assessment, followed with tests for neuro-dynamic flexibility, when appropriate. The preliminary systematic review (see Chapter 2), indicated that standard lower limb myotome, dermatome and tendon reflex tests, combined with either passive straight leg raise (PSLR) or femoral nerve stretch as indicated, improved the accuracy to diagnose radiculopathy in CLBP (Iversen et al., 2013). The McKenzie Institute Lumbar Spine Assessment (MII, 2005) includes these neurological and neuro-dynamic assessments, thus adding further support for including a complete McKenzie Institute Lumbar Spine Assessment within the construct of the final physiotherapy CLBP examination algorithm.

In addition to the McKenzie Institute Lumbar Spine Assessment (MII, 2005), the diagnostic value of specific individual tests and clustered examination procedures for discogenic pain (Laslett, Aprill, et al., 2006; Laslett, Öberg, et al., 2005), radiculopathy (Iversen et al., 2013) and FJS (Laslett, McDonald, et al., 2006; Laslett

et al., 2004) were identified from the preliminary systematic review (see Chapter 2). Subsequently, these were considered for inclusion in the CLBP examination algorithm.

The systematic reviews reported in Chapters 2 and 3 did not identify diagnostic value and reliability of any additional assessment processes for CLBP. Accordingly, for the final CLBP algorithm, it was necessary to choose additional examination processes according to their validity and reliability as reported from previous studies conducted in heterogeneous LBP populations.

Alqarni et al. (2011) conducted a systematic review that reported that passive accessory intervertebral movement (PAIVM) tests and the prone instability and posterior shear test, combined with other tests, assist LBP classification of SI. Combined movement tests, particularly lumbar extension combined with ipsilateral rotation to the side of the LBP, have also shown diagnostic value for FJS when combined with other tests (Hengeveld & Banks, 2014; Wilde et al., 2007). In addition, three or more positive responses on sacroiliac joint provocation tests of distraction, compression, thigh thrust, Gaenslen and sacral thrust, in the absence of CP, have shown sufficient diagnostic value for SIJS (Laslett, 2008; Laslett, Aprill, et al., 2005; Laslett et al., 2003; Peterson et al., 2017). Thus, based on available best evidence, the most appropriate components of the physical examination were chosen for inclusion in the final structure of the comprehensive CLBP physiotherapy examination algorithm.

4.3 Structure of the Algorithm

This section describes the structure and ordered flow of the comprehensive physiotherapy CLBP examination algorithm developed, which will subsequently be referred to as MK-C. Optimal structure of the algorithm required reproducible and

consistent flow of examination procedures, presented in a compact user friendly

form, with minimal risk of ambiguity in reporting patient response to the various

elements of each component. An MK-C assessment form was designed (see Figure

4.1), which considered these factors, and was supplementary to self-reported ODI

(see Appendix 1A), RM (see Appendix 1B), MSPQ (see Appendix 1C) and VAS

(see Appendix 1D), and a standard McKenzie Institute Lumbar Spine Assessment

(MII, 2005). The supplementary purpose-designed MK-C assessment form (see

Figure 4.1) included recorded measures of lumbar ROM, and a series of clinical tests

to identify secondary or co-existing patho-anatomical sources of symptoms as

previously described (see Chapter 4.2.1, 4.2.2 & 4.2.3).

4.4 MK-C Examination Procedures

This section details the method of examination and clinical reasoning process

for the MK-C. Some content of this section includes methods detailed in the

publication 'Classification Characteristics of a CLBP Population Using a Combined

Mckenzie and Patho-Anatomical Assessment' (Flavell, Gordon, & Marshman, 2016;

see Appendix 3).

A physiotherapist completed the MK-C in a consistent and sequential

manner. The ODI, RM, MSPQ and VAS were completed, followed by a standard

McKenzie Institute Lumbar Spine Assessment, prior to the physical examination

processes detailed in the MK-C supplementary assessment form (see Figure 4.1).

A McKenzie classification of derangement syndrome, dysfunction syndrome,

postural syndrome, mechanically inconclusive or 'other' was assigned immediately

following the McKenzie Institute Lumbar Spine Assessment, according to the

following definitions previously described by McKenzie and May (2003).

Participant number:

Complete in conjunction with the McKenzie Institute Lumbar Spine Assessment

Height	cm
Weight	kg
VAS	mm
Symptom duration	months
RM (24 questions, each marked as 1), scored as total/24.	/24
ODI (10 questions, each marked from 0 to 5), scored as (total/50) × 100.	%
MSPQ (13 questions each marked from 0 to 3), scored as total/39.	/39

Range of motion and neuro-dynamic measurements

Movement loss	Measure	Symptoms? Y/N
Flexion	cm	
Extension	cm	
Lateral flexion right	cm	
Lateral flexion left	cm	
Rotation right	cm	
Rotation left	cm	
PSLR right	°	
PSLR left	°	
Slump test	°	
Slump test	°	

Centralisation phenomenon? Y/N

Complete following the McKenzie assessment

FJS testing protocol

	Yes	No
Localised unilateral paraspinal pain at onset		
Unilateral pressure over facet joint at suspected level replicates or aggravates pain, +/- unilateral muscle spasm.		
Lack of radicular features		
Pain eased in flexion or sitting/Pain not worse with forward flexion.		
Referred pain never below knee joint		
Local unilateral passive movement shows reduced range of motion or increased stiffness ipsilateral to the pain.		
Painful lumbar spine extension.		
Painful lumbar spine extension, lateral flexion or rotation to the ipsilateral side/+ve ER test.		

Combined movement

Extension + LF/Rotn R	+ve −ve
Extension + LF/Rotn L	+ve −ve

PAIVM

Central L1	R P			
Central L2	R P			
Central L3	R P			
Central L4	R P			
Central L5	R P			
R unilateral T12/L1	R P	L unilateral T12/L1	R P
R unilateral L1/L2	R P	L unilateral L1/L2	R P
R unilateral L2/L3	R P	L unilateral L2/L3	R P
R unilateral L3/L4	R P	L unilateral L3/L4	R P
R unilateral L4/L5	R P	L unilateral L4/L5	R P
R unilateral L5/S1	R P	L unilateral L5/S1	R P

SIJS testing protocol

Provocation tests

Distraction	+ve −ve
Compression	+ve −ve
Gaenslan	+ve −ve
Thigh thrust	+ve −ve
Sacral thrust	+ve −ve
Patrick	+ve −ve

Recurrent locking, catching, giving way during active movements or apprehension (demonstrates anxiety about the sensation of collapse due to LBP during movement)	+ve −ve
Aberrant motion with trunk ROM, such as hand-thigh walking on extension or instability catch sign test (inability to bend forwards and return to an erect position due to sudden onset LBP)	+ve −ve
Intervertebral motion testing to determine hypermobility Spinal level=	+ve −ve
Prone instability test	+ve −ve
Posterior shear test	+ve −ve

Classification:

Spinal level:

Figure 4.1. MK-C supplementary assessment form.

Participants with the following characteristics were classified with

derangement syndrome.

History included:

- local and/or referred pain that was constant or intermittent

- symptoms that were episodic in nature, variable over time, with an acute
  or gradual onset

- symptom severity and distribution that were aggravated or improved by
  certain postures or movements.

Physical examination included:

- reduced or 'blocked' ROM

- repeated movement testing centralised or peripheralised pain and
  increased or decreased ROM. (Pain centralisation has been described
  elsewhere [May and Aina, 2012].)

Participants with derangement syndrome were sub-classified as central symmetrical,
asymmetrical above knee or asymmetrical below knee (McKenzie & May, 2003).

Participants with the following characteristics were classified as **postural
syndrome.**

History included:

- younger age

- sedentary lifestyle

- time-dependent onset of intermittent local symptoms due to prolonged
  postural loading of healthy tissues

- symptoms commonly produced with slumped sitting, never produced by
  movement or activity.

Physical examination included:

- poor spinal posture

- symptom abolishment by postural correction

- pain production or abolishment by static testing

- unaffected ROM

- no effect with repeated movement testing.

Participants with the following characteristics were classified as **dysfunction syndrome.**

History included:

- previous derangement syndrome or traumatic injury, persistent poor posture or spinal degeneration

- intermittent symptoms that abated and did not persist once loading stopped

- local pain, unless symptoms referred into the lower limb because of adherence of at least one lumbo-sacral NR, otherwise classified as adherent nerve root (ANR; McKenzie & May, 2003).

Physical examination included:

- pain production that was consistent with a specific direction and ROM

- reduced ROM in one or more anatomical planes

- symptoms that were produced but never worsened by specific repeated movements.

Participants were classified as **inconclusive** if symptoms were of unknown lumbar joint pathology and were affected by movement or posture, but with variable loading response, and no one syndrome was predominant on first assessment (McKenzie & May, 2003).

Participants classified as '**other**' included those with suspected chronic pain states or with symptoms unrelated to the lumbar spine or sacroiliac joint, and those who did not fit into one of the previously described syndromes (McKenzie & May, 2003). McKenzie and May (2003) described several presentations as 'other', including suspected serious spinal pathology and the presence of 'red flags'. However, as these presentations were exclusion criteria for this study, 'other' was not represented.

To conclude the examination process, immediately following the McKenzie Institute Lumbar Spine Assessment, the physiotherapist conducted the additional MK-C clinical tests using the standardised MK-C supplementary assessment form (see Figure 4.1). A schematic representation of the MK-C algorithm is presented in Figure 4.2. The MK-C consisted of physical examination processes and clustered patho-anatomical clinical tests conducted in the following order:

- combined lumbar movements extension, ipsilateral rotation and lateral flexion conducted bilaterally (Hengeveld & Banks, 2014)

- central and unilateral PAIVM T10 to L5/S1 (Hengeveld & Banks, 2014)

- nine clinical indicators of FJS (Laslett, McDonald, et al., 2006; Wilde et al., 2007; see Table 4.1)

- six clinical indicators of SIJS (Laslett, Aprill, et al., 2005; Laslett et al., 2003; see Table 4.2)

- five clinical indicators of lumbar spine instability (Abbott et al., 2005; Fritz et al., 2005; Kotilainen & Valtonen, 1993; see Table 4.3).

Three of the 12 clinical indicators for FJS identified by Wilde et al. (2007) were not included in the algorithm because they were either not relevant to a physical

examination or were imaging results to which for this study the physiotherapist was blinded. The excluded indicators were:

- positive response to intra-articular facet joint injection
- pain relieved by fluoroscopically guided double anaesthetic blocks of the medial branch of the dorsal ramus supplying the facet joint
- radiography unreliable and non-diagnostic for facet joint pain.

Sacroiliac joint provocation tests were conducted in a consistent order of distraction, compression, thigh thrust, and Gaeslen's and sacral thrust according to methods described previously (Laslett, Aprill, et al., 2005; Laslett et al., 2003). A schematic representation of the MK-C algorithm is presented in Figure 4.2.

On completion of the MK-C, participants were assigned to one of nine final classifications—derangement syndrome, postural syndrome, dysfunction syndrome, inconclusive, discogenic, FJS, SIJS, SI and mixed—according to history, clinical examination responses and operational definitions previously reported (Abbott et al., 2005; Fritz et al., 2005; Kotilainen & Valtonen, 1993; Laslett, 2008; Laslett, Aprill, et al., 2005; Laslett, McDonald, et al., 2006; Laslett, Öberg, et al., 2005; Laslett et al., 2003; McKenzie& May, 2003; Wilde et al., 2007). A summary description is detailed in the following paragraphs.

Discogenic: This classification comprised participants classified as derangement syndrome with the McKenzie Institute Lumbar Spine Assessment, who demonstrated DP and presented with CP combined with reduced lumbar extension or a feeling of vulnerability on flexion or both, and demonstrated no indicators for SI with the MK-C. This combination of factors, in the presence of persistent pain, is predictive of discogenic pathology with a reported +ve LR (95% CI) of 6.7 [0.95, 50.0] and −ve LR of 0.73 [0.61, 0.97] (Laslett, Aprill, et al., 2006).

86
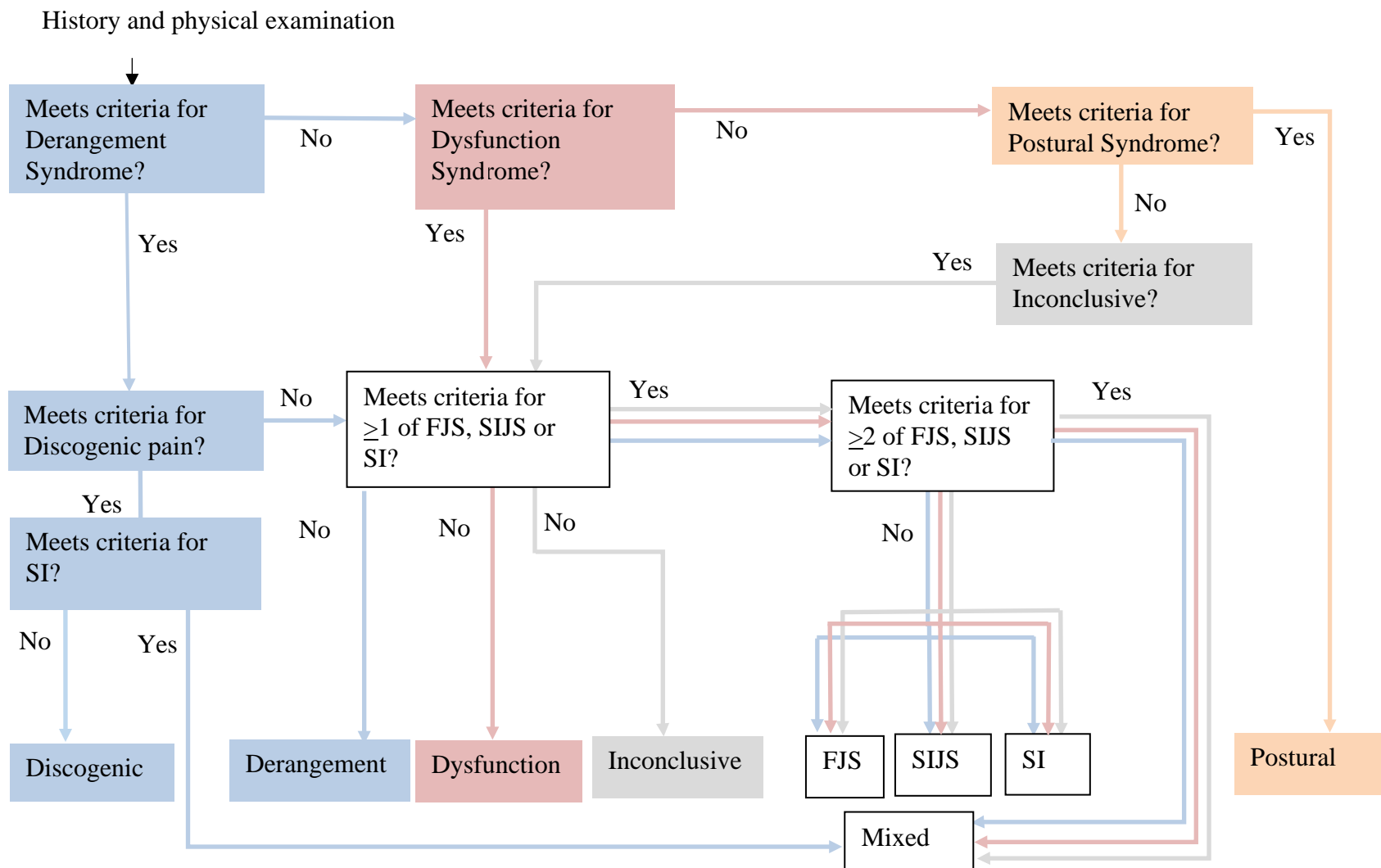
Figure 4.2. MK-C examination algorithm. FJS = facet joint syndrome; SI = spinal instability; SIJS = sacroiliac joint syndrome.

Table 4.1

MK-C Clinical Indicators for Facet Joint Syndrome[a]

Clinical indicator		MK-C examination component.
1	Absence of centralisation.	McKenzie Institute Lumbar Spine Assessment.
2	Localised unilateral low back pain/paraspinal pain.	Body chart.
3	Unilateral pressure over facet joint at suspected level replicates or aggravates pain, +/- unilateral muscle spasm.	Passive accessory intervertebral movement of lumbar spine.
4	Lack of radicular features.	Body chart.
5	Pain eased in flexion or sitting/Pain not worse with forward flexion.	McKenzie Institute Lumbar Spine Assessment.
6	Pain if referred to the leg is above the knee.	Body chart.
7	Local unilateral passive movement shows reduced range of motion or increased stiffness ipsilateral to the pain.	Passive accessory intervertebral movement of lumbar spine.
8	Painful lumbar spine extension.	McKenzie Institute Lumbar Spine Assessment.
9	Painful lumbar spine extension, lateral flexion or rotation to the ipsilateral side/+ve extension rotation test.	Combined movement testing.

Note. Source: Hengeveld and Banks (2014); Laslett, McDonald, et al. (2006); and Wilde et al. (2007).
[a]Minimum criteria required 1 plus 2,5,6,& 9.

Table 4.2

MK-C Clinical Indicators for Sacroiliac Joint Syndrome[a]

Clinical indicator	MK-C examination component immediately following tests for facet joint syndrome
1 Absence of centralisation	McKenzie Institute Lumbar Spine Assessment.
2 +ve Distraction test	
3 +ve Compression test	
4 +ve Thigh thrust	Sacroiliac Joint provocation tests.
5 +ve Gaenslen test	
6 +ve Sacral thrust	

Note. Source: Laslett (2008), Laslett et al. (2005), and Laslett et al. (2003).

[a]Minimum criteria required 1 plus >3 Sacroiliac Joint provocation tests.

Derangement: This classification comprised participants who demonstrated DP and thus were classified as derangement syndrome with the McKenzie Institute Lumbar Spine Assessment, but demonstrated no CP to indicate discogenic pain and no other patho-anatomical source of CLBP because of the absence of clinical indicators for FJS, SIJS and SI on MK-C examination.

Dysfunction: This classification comprised participants who were classified as dysfunction syndrome with the McKenzie Institute Lumbar Spine Assessment and demonstrated no patho-anatomical source of CLBP because of the absence of clinical indicators for FJS, SIJS and SI on MK-C examination. Dysfunction with ANR was classified in participants identified as dysfunction syndrome, with additional criteria of recent sciatica and intermittent leg pain produced on movement (McKenzie & May, 2003).

Inconclusive: This classification comprised participants classified as inconclusive with the McKenzie Institute Lumbar Spine Assessment, who demonstrated no indicators of FJS, SIJS or SI with the MK-C.

Facet joint syndrome: This classification comprised participants classified as dysfunction syndrome, inconclusive or derangement syndrome with the McKenzie Institute Lumbar Spine Assessment, who demonstrated no CP indicating discogenic pain, who had an MSPQ score greater than 13 and who tested positive to elements 2, 5, 6 and 9 of the MK-C FJS cluster (see Table 4.1). This cluster represents a previously identified CPR with a +ve LR 7.6 [4.5, 13.7] and a −ve LR 0.0 [0.0, 0.35] (Laslett, McDonald, et al., 2006).

Sacroiliac joint syndrome: This classification comprised participants classified as dysfunction syndrome, inconclusive or derangement syndrome with the McKenzie Institute Lumbar Spine Assessment, who demonstrated no CP indicating

discogenic pain and whose symptoms were elicited or increased with three or more of the MK-C-included SIJ provocation tests (Laslett, 2008; Laslett, Aprill, et al., 2005; Laslett et al., 2003; see Table 4.2). Previously this CPR has yielded a +ve LR of 6.97 [2.39, 20] and a −ve LR of 0.10 [0.02, 0.68] (Laslett, 2008).

Spinal instability: Because of the equivocal validity and very small to moderate +ve LR reported for clinical signs of lumbar spine instability (Alqarni et al., 2011), this classification was assigned in participants classified as dysfunction syndrome, inconclusive or derangement syndrome on the McKenzie Institute Lumbar Spine Assessment, who demonstrated no CP indicating discogenic pain when all five MK-C SI indicators were present (see Table 4.3). Indicators 1 and 2 were assessed during the McKenzie Institute Lumbar Spine Assessment. Indicator 3 (+LR 2.4 [0.9; 6.4] and -LR 0.7 [0.4, 1.0]), indicator 4 (+LR 1.4 [0.8, 2.5] and -LR .7 [0.4, 1.2]) and indicator 5 (+LR 1.1 [0.7, 1.8] and -LR 0.9 [0.5, 1.5]; Alqarni et al., 2011) were assessed using methods previously described (Abbott et al., 2005; Fritz et al., 2005; Kotilainen & Valtonen, 1993).

Mixed: This classification comprised participants classified as derangement syndrome on the McKenzie Institute Lumbar Spine Assessment and with one of the following two criteria:

- with indicators of discogenic pain and positive responses for all indicators of SI on the MK-C

- without indicators of dysfunction syndrome, inconclusive or CP to indicate discogenic pain, but with demonstrated positive responses for two or more test clusters for FJS, SIJS or SI, where no one of these classifications predominated on the MK-C.

Postural syndrome: On the basis of McKenzie's definition of postural syndrome (McKenzie & May, 2003), further examination for patho-anatomical sources of symptoms was not indicated following the McKenzie Institute Lumbar Spine Assessment, and the final classification for these participants remained as postural syndrome.

4.5 Piloting of the MK-C

Piloting of the MK-C was conducted at the Townsville Hospital study site, in the examination room of the CLBP specialist clinic. Piloting was conducted during one day, with consenting patients who attended the neurosurgeon's once weekly CLBP clinic. This provided the opportunity to appraise the examination process for usability, and to assess how the physiotherapy examination and other aspects of the study data collection would integrate into the processes, protocols and logistics of the Townsville Hospital CLBP clinic. Particular aspects under evaluation were suitability of methods to identify potential participants, maintenance of study blinding, and that the usual patient-centred clinic processes and timing were not compromised by the research data collection protocol.

Six participants were recruited and underwent the full MK-C examination. No modifications to the examination process were identified. Minor study recruitment, issues related to administration of the invitations to the study were identified and resolved via discussion with clinic personnel. Methods of data concealment were evaluated without requiring further modification.

Table 4.3

Clinical Indicators of Spinal Instability

Clinical indicator	MK-C examination component immediately following tests for facet and sacroiliac joint syndrome
1 Reports of recurrent locking, catching, giving way during active movements or apprehension (demonstrates anxiety about the sensation of collapse due to low back pain during movement).	McKenzie Institute Lumbar Spine Assessment movement tests.
2 Aberrant motion with trunk ROM, such as hand-thigh walking on extension or instability catch sign test (inability to bend forwards and return to an erect position due to sudden onset low back pain).	McKenzie Institute Lumbar Spine Assessment movement tests.
3 Intervertebral motion testing to determine hypermobility.	Passive accessory intervertebral movements of lumbar spine.
4 Prone instability test.	Sacroiliac joint provocation tests.
5 Posterior shear test.	

Note. ROM = range of movement. Source: Abbott et al. (2005); Fritz et al. (2005); and Kotilainen and Valtonen (1993).

4.6 How This Chapter Informed Subsequent Stages of the Research

This chapter detailed the reasoning and evidence that underpinned the MK-C examination components, defined the classification categories, and explained the examination procedure and clinical reasoning used to reach each final classification of the MK-C. Additionally, successful piloting of the MK-C in the relevant clinical setting supported its suitability for the CLBP studies subsequently reported (see Chapters 6 & 7). However, prior to conducting these CLBP studies, the intra-examiner reliability of the MK-C-included lumbar spine ROM and neuro-dynamic tests needed to be established. This is reported in Chapter 5 of this thesis.

Chapter 5: Reliability of Range of Motion and Neuro-Dynamic Measurement

This chapter supports the content of Chapters 2, 3 and 4, and addresses the second research question (see Chapter 1): 'Are the physical measurements typically used by physiotherapists in LBP examination reliable to assess an asymptomatic population matched to, and typical of CLBP patients?' Particularly, this chapter reports intra-examiner reliability of lumbar spine and neuro-dynamic flexibility measurements. The study was conducted to establish the intra-examiner reliability of the physiotherapist to conduct the MK-C-included lumbar spine ROM and neuro-dynamic test measurements in a healthy population with characteristics of BMI and age that simulate CLBP patients, to indicate their suitability for inclusion in the CLBP examination algorithm.

This chapter reports a study entitled 'Intra-Examiner Reliability of Lumbar Spine and Neuro-Dynamic Flexibility Measurements in an Older and Overweight Healthy Asymptomatic Population' (Flavell, Gordon, & Watt, 2017; see Appendix 3).

5.1 Introduction

Health professionals such as physiotherapists use spinal movement and neuro-dynamic flexibility tests as measures of intervention effect. Such movement measurements constitute part of an overall assessment process for patients with LBP; however, according to previous research, reliability of lumbar spine movement tests is dependent on many factors, including the type of population being measured (Doriot & Wang, 2006). There is potential for increased measurement error due to greater soft-tissue excursion relative to underlying bony landmarks in older people (Kuo, Tully, & Galea, 2008), and excessive adipose tissue relative to underlying

bony landmarks in those with high BMI (Broadbent et al., 2000). Moreover, clinicians should consider the evidence that suggests a relationship between obesity and LBP (Cimolin et al., 2011; Shiri et al., 2010) and the importance of being able to reliably assess spinal movement in obese people with CLBP.

Measurement with tape measures and goniometry is used clinically by many health professionals, particularly physiotherapists. Clinicians and researchers use these simple tools to measure spinal movement and neuro-dynamic flexibility, predominantly to provide an initial measure and to subsequently evaluate change via repeated measures following an intervention. The degree of error in these measurements when repeated over time should be minimal, otherwise the results will not represent true change. Reliability reports for lumbar spine and neuro-dynamic movement tests have varied. Beattie, Rothstein, and Lamb (1987) evaluated the modified Schöber method for lumbar extension measurement, and Leard, Crane and Ball (2009) the Schöber method for lumbar flexion measurement (Macrae & Wright, 1969). These studies reported good intra-examiner reliability for lumbar extension and flexion measurement. However, Frost, Stuckey, Smalley and Dorman (1982) reported good reliability only for lumbar flexion and lateral flexion, and poor reliability for extension and rotation measurements. In comparison with Beattie et al. (1987), these researchers used an alternative method, which resulted in combined thoraco-lumbar extension measurement, not specific to measurement of lumbar movement. In addition, despite using different measurement methods, both Frost et al. (1982) using a tape measure to plinth method, and Leard et al. (2009) using inclinometry, reported poor reliability for hip flexion during the PSLR, and one study reported good reliability when measuring knee extension during the slump test (Tucker, Reid, & McNair, 2007).

Various methods for measuring lumbar ROM with a tape measure have been described (Gill, Krag, Johnson, Haugh, & Pope, 1988; Hyytiäinen, Salminen, Suvitie, Wickström, & Pentti, 1991; Macrae & Wright, 1969; Portek, Pearcy, Reader, & Mowat, 1983). Specifically, Hyytiäinen et al. (1991) showed that the modified Schöber method (Beattie et al., 1987) to assess lumbar sagittal movements in normal subjects had greater accuracy and higher intra-examiner reliability for measuring lumbar flexion than did the Schöber method (Macrae & Wright, 1969). Further, the modified Schöber method has been validated by good correlation with radiographic measures (Portek et al., 1983), and has greater reliability for assessment of flexion and extension when compared with a two-inclinometer method (Gill et al., 1988). Good intra-examiner reliability has been reported for the finger-to-floor method to measure lateral flexion (Frost et al., 1982). In contrast, use of goniometry for hip joint measurement during neuro-dynamic testing using the PSLR (Leard et al., 2009) has been reported to have low intra-examiner reliability. However, measurement of knee extension with electro-goniometry has reported good reliability during the slump test (Tucker et al., 2007). Nevertheless, electro-goniometers are more difficult to use than simple tools such as the universal goniometer and may not be readily available in the clinical setting. Tape measures and universal goniometers remain simple, user friendly and frequently available clinical tools for joint measurement. While the results of previous reliability studies are useful, all of them were conducted with participants younger than 50 years and with normal BMI (20–24; Beattie et al., 1987; Frost et al., 1982; Hyytiäinen et al., 1991; Leard et al., 2009; Tucker et al., 2007). Therefore, the results are not specific to an older or overweight population.

Using standard protocols and equipment, this pilot study investigated the intra-examiner reliability of lumbar spine ROM and neuro-dynamic flexibility in a healthy asymptomatic population of overweight, older adults selected to represent the demography of the CLBP population.

5.2 Materials and Methods

5.2.1 Study design

Blinded test–retest intra-examiner single-group reliability study.

5.2.2 Ethics

All participants received a verbal and written description of the study. Written consent was obtained prior to data collection, and the study was conducted with approval of the institutional Human Ethics Research Committee (H4547; see Appendix 4a).

5.2.3 Recruitment

Recruitment was by email, word of mouth, 'flyers' distributed within the local community and attendance at a community meeting.

5.2.4 Participants

Participants aged 50 years or over with a BMI of 24 or greater were recruited to the study. Participants were excluded in the presence of any musculoskeletal or medical condition that would preclude them from safely performing any part of the testing protocol.

5.2.5 Experimental procedures

5.2.5.1 Experimental equipment and measurements

Standard protocols were used to measure height with a stadiometer, weight with weighing scales (Norton & Olds, 2006), lumbar spine flexion, extension, lateral

flexion and rotation with a tape measure, and hip and knee flexion with a goniometer (Clarkson, 2005).

Six lumbar spine movements were measured: flexion; extension; right and left lateral flexion; and right and left rotation. In addition, to assess neuro-dynamic flexibility, range of hip flexion during the PSLR (Goddard & Reid, 1965) and knee extension during the slump test (Maitland, 1986) were measured.

5.2.5.2 Examiner

The examiner, a physiotherapist with 27 years of clinical experience, conducted all examination tests and measurements. An assistant recorder, experienced in reading a goniometer and standard tape measure, recorded each measurement.

5.2.5.3 Participants

Participants consented to remove upper body outer garments and shoes, and wore shorts, during all procedures. Participants were advised and instructed on the study procedures and given the opportunity to warm up with simple active movements prior to testing. To avoid any possible risk of injury or discomfort during testing, participants were instructed to perform the test movements to their maximum ability but not 'over stretch' into the movements.

5.2.5.4 Examination tests

Participants completed eight examination tests using standard measurement protocols (Beattie et al., 1987; Clarkson, 2005): six lumbar spine movement tests measured with the tape measure and two neuro-dynamic tests measured with the goniometer. To minimise bias and blind the clinician, the numbers on the face of the measurement devices were covered to the clinician. The assistant was able to view the numbers and recorded each measurement on a standardised record sheet. All

measurements were completed once, then participants rested for 15 min before the movements were repeated and remeasured. This procedure was repeated five times. All tests were recorded during one 90-minute attendance session.

5.2.5.5 Lumbar flexion and extension

The skin was marked according to the methods described by Beattie et al. for the modified Schöber method as per Figure 5.1. For lumbar flexion, participants were asked to bend forwards from a standing position, sliding both hands along the anterior aspect of both thighs towards the floor. For lumbar extension, with hands resting on iliac crests, participants were asked to bend backwards from a standing position. For both tests, once the participants had reached their maximum ROM, the tape measure was placed along a line between landmarks C and B and the distance measured (Plate 5.1). The resultant difference between the initial (15 cm) and final measurements was calculated to determine ROM.

5.2.5.6 Right and left trunk lateral flexion

Participants were asked to laterally flex to the right and reach towards the floor while sliding their ipsilateral hand with fingers extended along the lateral aspect of their lower limb. Upon reaching their maximum ROM the examiner used a standard tape measure to record the distance between the tip of the third digit and a point directly inferior to this on the floor. The procedure was then conducted to the contralateral side.

5.2.5.7 Right and left trunk rotation

On both the left and the right side, the skin overlying the midpoint of the lateral border of the acromion and the greater trochanter was identified and marked. With arms crossed in front of their chest, participants were asked to turn to the right while keeping their feet fixed on the ground. Upon reaching their maximum ROM,

the distance between the markers on the acromion and the greater trochanter was

recorded. The procedure was then conducted to the left side.
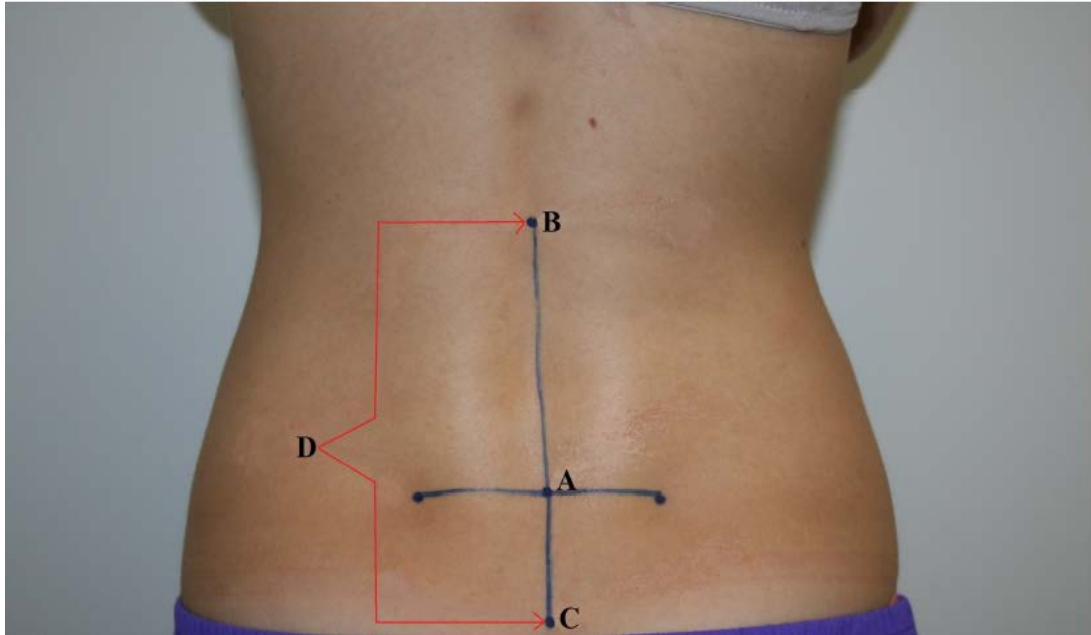
Plate 5.1. Illustration of skin landmarks for the modified Schöber test: (A) 0 cm

mark on the spinous process of S2, identified by a line joining both posterior superior

iliac spines; (B) 10 cm superior to A; (C) 5 cm inferior to A; (D) 15 cm initial

measurement distance.

5.2.5.8 Unilateral right slump

Participants were seated in a standardised position on the plinth with their

hands placed on their anterior thigh and lower legs hanging freely. The middle of the

lateral knee joint line on the participants' right side was identified and marked with

removable tape. The examiner then conducted a slump test procedure (Maitland,

1986) until participants' maximum active right knee extension was achieved. The

right knee joint angle was then measured with the goniometer using a standard

protocol (Clarkson, 2005).

5.2.5.9 Unilateral left passive straight leg raise

Participants rested in a supine position on a treatment plinth without pillows. The participants' left greater trochanter of the femur was identified and marked with removable tape. The examiner then conducted a PSLR procedure (Goddard & Reid, 1965) by passively raising the left lower limb off the plinth while ensuring that the right lower limb remained still with no pelvic tilting. The left knee joint was maintained in an extended position. Maximum PSLR was reached when the participant reported 'maximum' stretch or any discomfort, or the examiner detected firm resistance or pelvic tilting. Participants' left hip flexion was then measured with the goniometer (Clarkson, 2005).

5.2.6 Statistical analysis

Descriptive statistics including the mean and standard deviation (*SD*) were calculated for height, weight, age and BMI. For each test, intra-class correlation coefficient (*ICC*), 95% CI, and standard error of measurement (SEM; Hopkins, 2000) were calculated. Statistical analysis was performed using the Statistical Package for the Social Sciences, version 20 (SPSS Inc., Chicago, IL).

5.3 Results

Nineteen volunteers (8 male and 11 female subjects) participated in the study (see Table 5.1). The *ICC*s were nearly perfect (*ICC* $\geq$ .90; Hopkins, 2000) for right (*ICC* = .90) and left (*ICC* = .94) lateral flexion, right (*ICC* = .93) and left (*ICC* = .99) rotation, and the slump test (*ICC* = .94). Very large correlations were reported for the PSLR (*ICC* = .87) and for flexion (*ICC* = .88) measurements, and a large correlation for extension (*ICC* = .68; see Table 5.2).

Table 5.1

Participant Demographics (N = 19)

	$\bar{x}$	SD	Range
Age (yr)	56.00	7.62	50–81
Weight (kg)	79.26	14.10	56.90–105.90
Height (m)	1.67	0.13	1.51–1.92
Body mass index	28.32	3.58	24.07–37.08

Table 5.2

Intra-Examiner Reliability

	ICC	95% CI	*SEM*
Flexion	.88	[0.78, 0.95]	0.36
Extension	.68	[0.49, 0.84]	0.58
Right lateral flexion	.90	[0.82, 0.96]	0.33
Left lateral flexion	.94	[0.88, 0.97]	0.27
Right rotation	.93	[0.87, 0.97]	0.27
Left rotation	.99	[0.98, 0.99]	0.12
Right slump	.94	[0.89, 0.98]	0.25
Left PSLR	.87	[0.78, 0.95]	0.37

Note. CI = confidence interval; *ICC* = intra-class correlation coefficient; PSLR = passive straight leg raise test; *SEM* = standard error of measurement.

5.4 Discussion

This study was conducted to evaluate the intra-examiner reliability of lumbar spine motion and joint measurements for evaluating neuro-dynamic flexibility in an older, overweight population. The study demonstrated that, across all tests, the intra-examiner reliability of the measurements for lumbar spine ROM and neuro-dynamic flexibility was nearly perfect. The lateral flexion, rotation and slump tests had the highest level of reliability and extension the lowest.

Older participants with higher BMI were recruited to this study to replicate the reported association between BMI, age and symptoms of CLBP. However, for ethical reasons, only asymptomatic volunteers were recruited because the potential

for repeated movement testing to provoke a pain response in symptomatic participants was high. The aim was to report the reliability of clinical measurement in the presence of the CLBP-associated factors of higher BMI and age.

For lumbar flexion and extension measurements, the study showed that reliability of the modified Schöber method (Beattie et al., 1987) was very high. This supports previous studies that have reported high reliability using similar methods (Frost et al., 1982; Gill et al., 1988; Hyytiäinen et al., 1991; Macrae & Wright, 1969; Portek et al., 1983).

Although lumbar extension measurement in this study showed high reliability ($ICC = .68$), it was the lowest level of reliability compared with the other measurements evaluated in this study. Interestingly, poor reliability of the modified Schöber method for extension has been reported previously, being related to the small amount of excursion available during this movement (Frost et al., 1982). In contrast, Beattie et al. (1987) reported ICCs of .90 and greater when measuring lumbar extension, in both symptom-free and symptomatic participants with 'significant limiting LBP'. In summary, the reliability reported for lumbar extension measurement using the modified Schöber method in the current study, combined with the findings of Beattie et al. (1987) and Frost et al. (1982), remains conflicting.

Nearly perfect ICCs were observed for the intra-examiner measurements of lateral flexion and rotation in this study. Frost et al. (1982) reported good intra-examiner reliability for the third finger-to-floor method of lateral flexion. Hence, the results of this study for reliability of lateral flexion support the findings of Frost et al. (1982). Conversely, the present study results did not support the poor reliability for rotation measurement reported by Frost et al. (1982).

Repeated measurements of the slump test and PSLR for neuro-dynamic flexibility showed nearly perfect and very high correlation, respectively. Previous studies have reported poor reliability when measuring hip flexion angle during the PSLR (Leard et al., 2009). Although few intra-examiner reliability studies have measured joint position during the PSLR and slump test, one study identified that pelvic position affected hip ROM during the PSLR (Cameron, Bohannon, & Owen, 1994). In this study, pelvic position was controlled as much as possible but, ultimately, reliability of this measurement might have been affected by difficulties in positioning the goniometer while maintaining lower limb position during the test. This was particularly challenging with the individuals in this study whose higher BMI might have reduced the accuracy of greater trochanter identification. The paucity of past research into measurement reliability for the PSLR prevents comparison.

A limitation to this study was the possible implication of progressive increases in ROM over time due to the visco-elastic properties of the spinal soft tissues. No research has previously reported the optimal time interval for repeated measures of the lumbar spine. In an effort to reduce confounding soft-tissue adaptations and on the basis of clinical judgement, the clinician set a minimum 15-minute interval between participant measurements. The reliability results of this study support the use of this time interval between measurements to reduce the effect of soft-tissue adaptation.

5.5 Conclusion

This research established levels of intra-examiner reliability of lumbar spine and neuro-dynamic flexibility measurements in an older and overweight, healthy asymptomatic population. The findings support the suitability of these methods and

tools for measurement in the clinical setting, and suggest that age and increased BMI do not adversely affect repeated measurement reliability. Nevertheless, the paucity of previous literature disallows comparisons with other population demographics for some measures. In particular, future studies should aim to evaluate whether symptomatic CLBP patients have a clinically significant effect on measurement reliability. Particularly, the neuro-dynamic measurements of the PSLR and slump test.

Key points:

- The results of this study indicated that the physiotherapist was highly reliable when measuring ROM and neuro-dynamic flexibility in an overweight, older population.

- High levels of reliability support the inclusion of these methods of measurement as part of the physiotherapy examination algorithm, which will be conducted for a typically older, overweight CLBP population in the study described in Chapter 6 of this thesis.

5.6 How This Chapter Informed Subsequent Stages of the Research

It was important to this research that, when using the physiotherapy CLBP examination algorithm, the physiotherapist's measurement of lumbar ROM and neuro-dynamic flexibility demonstrated high reliability. This study explained the methods and reported the intra-examiner reliability of lumbar spine ROM and joint ROM associated with neuro-dynamic testing. The results supported the inclusion of these measurements in the proposed MK-C examination algorithm.

Chapter 6: Classification Characteristics of a Secondary-Care Chronic Low Back Pain Population

This chapter addresses the third research question (see Chapter 1): 'What are the classification characteristics of a CLBP population using an evidence-based physiotherapy examination algorithm specifically designed for CLBP, including demographics, the Oswestry Disability Index (ODI), Roland–Morris disability questionnaire (RM), Modified Somatic Pain Perceptions Questionnaire (MSPQ), symptom duration, and 11-point visual analogue scale (VAS)?'

The content of this chapter includes sections of the publication 'Classification Characteristics of a Chronic Low Back Pain Population Using a Combined Mckenzie and Patho-Anatomical Assessment' (Flavell, Gordon, & Marshman, 2016; see Appendix 3).

6.1 Introduction

There is evidence to suggest that some LBP characteristics differ in CLBP. For example, a higher proportion of female patients (DePalma et al., 2012; Viniol et al., 2013) and increased BMI has been associated with CLBP (Cimolin et al., 2011; Heuch et al., 2010). Increased peripheral symptom distribution has also been associated with the longer duration and increased intensity of symptoms reported in CLBP (Prins et al., 2013). Furthermore, the increased age evident in CLBP raises the probability of FJS and SIJS (DePalma, Ketchum, & Saullo, 2011). Notably, modified somatic pain perceptions are also often a feature of chronic pain conditions (Ardic & Toraman, 2002). Notwithstanding, a comprehensive report of demographic, functional and symptom characteristics has been conducted in only one previous LBP study (Eirikstoft & Kongsted, 2014), and no studies have reported and compared the classification characteristics of a CLBP population.

Accordingly, the objectives of this study were to report characteristics and classification frequencies of a CLBP population examined using the MK-C, and report on classification differences in CLBP characteristics of BMI, age, gender, symptom duration and intensity, disability using the ODI and RM, and pain somatisation using the MSPQ.

6.2 Method

6.2.1 Study design

A prospective cross-sectional study was registered with the Australian New Zealand Clinical Trials Registry on 6 March 2013 (ACTRN: 12613000267752). This study was conducted by a McKenzie Institute Lumbar Spine Assessment–trained physiotherapist during the weekly CLBP clinic of a specialist neurosurgeon at the Townsville Hospital, between July 2012 and March 2014. Patients who attended the clinic resided in outer regional, rural and remote Australian communities and were referred from both primary and secondary healthcare sources.

6.2.2 Ethics

Study approval was obtained from the Human Research Ethics Committees of James Cook University and Townsville Hospital (JCUH4387/HREC10QTHS53; see Appendix 4B). Participants received a detailed explanation of the study and provided written consent prior to commencement of the study.

6.2.3 Participants and examiner

A convenience sample of consecutive patients who attended the weekly specialist CLBP clinic was recruited to the study. A McKenzie-trained physiotherapist, with more than 15 years of experience using the McKenzie Institute Lumbar Spine Assessment and almost 30 years of experience in LBP examination, screened volunteers for inclusion and exclusion criteria, and conducted all

examination procedures. The physiotherapist was blinded to medical records and imaging results for the purposes of this study.

6.2.4 Inclusion criteria

Participants were included who were 18 years of age or older and who presented with CLBP, defined as pain between the level of the 12th thoracic vertebra and the buttock crease, with or without associated lower limb symptoms (Jones et al., 2003) that had persisted for longer than 12 weeks (Reneman et al., 2006) and was adversely influencing functional status.

6.2.5 Exclusion criteria

Exclusion criteria were previous surgery of the lumbo-sacral spine; inability to tolerate the physical examination; presence of medical 'red flags' indicative of potentially serious medical conditions; pregnancy; inability to communicate; progressive neurological disturbance; current litigation, insurance or other compensation claims; previous lumbar spine infection, tumour, fracture or osteoporosis; and medical conditions that result in CLBP, such as fibromyalgia and ankylosing spondylitis.

6.2.6 Examination procedure and data collection

The data collection and examination process have been described in the preceding chapter of this thesis (see Chapter 4).

6.2.7 Data management

Baseline ODI, RM and MSPQ were converted to a percentage of raw score, and BMI as a measure of participant's relative size was calculated from height and weight measurements as kg/m^2 (Dahl, Fauth, Ernsth-Bravell, & Hassing, 2013). All data were imported to SPSS statistical analysis software, version 22 (SPSS Inc., Chicago, IL).

Descriptive statistics were generated for CLBP demographics and characteristics. Numerical variables were analysed for normality of distribution and reported using mean and SD ($\overline{x} \pm SD$), or median (*Mdn*) with interquartile range (*IQR*). MK-C classification categories were reported as proportions (%). Bivariate statistical tests were conducted as appropriate to assess for significant differences ($p < .05$) in demographic and CLBP characteristics among MK-C classification categories. The dysfunction syndrome and dysfunction ANR classification were pooled for statistical analysis purposes. All MK-C data were complete and no indeterminate results existed. Classification categories with less than five participants were reported in this study, but no statistical analysis was conducted as numbers were considered too low for statistical inference.

6.3 Results

Between July 2012 and March 2014, all patients who attended the CLBP clinic ($N = 316$) were invited to the study. The study recruited 62% ($N = 197$) of invited participants. Following application of inclusion and exclusion criteria, 76% ($N = 150$) of the 197 volunteers were admitted to the study. Reasons for exclusion included previous surgery of the lumbo-sacral spine ($N = 9$); current litigation, insurance or other compensation claims ($N = 4$); non-lumbar spinal pain ($N = 6$); ankylosing spondylitis or systemic lupus erythematosus ($N = 2$); and inability to tolerate the physical examination because of poor standing balance or other limiting co-morbidities ($N = 26$). Table 6.1. summarises the results of the MK-C classification.

Table 6.1

Results of the MK-C Classification

	$N = 150$	%
Discogenic	35	23
Derangement syndrome	4	3
Dysfunction syndrome	12	8
Inconclusive	9	6
Facet joint syndrome	74	49
Sacroiliac joint syndrome	1	1
Spinal instability	6	4
Mixed	7	5
Postural syndrome	2	1

6.3.1 Demographic characteristics

Descriptive statistics for the characteristics of the study population are summarised in Table 6.2. Male subjects represented 35% and female subjects 65% of participants. Male subjects reported significantly lower MSPQ ($p = .04$, CI [−10.45, −0.4]), and ODI ($p = .03$, CI [−10.28, 0.46]) scores than did female subjects. No other significant differences existed between genders for any descriptive factor.

6.3.2 Classification profile using the MK-C

Figure 6.1 shows the final classification frequencies derived following the initial McKenzie Institute Lumbar Spine Assessment component of the MK-C examination algorithm. Participant examination most frequently resulted in an MK-C classification of FJS (49%) and least frequently SIJS (1%). Of all participants identified as derangement syndrome (32%) following the initial McKenzie classification component of the MK-C, eight were negative for the intervertebral disc as the source of symptoms. Of these cases, 50% were negative for additional patho-anatomical sources of their symptoms resulting in a final MK-C classification of

derangement syndrome (see Figure 6.1). The remaining 50% tested negative for SIJS and SI, but positive for FJS, and therefore were classified as such.

Discogenic pain indicated by the presence of CP was evident in 83% of participants identified with derangement syndrome on initial McKenzie classification. Of these cases, four also tested positive for SI with the MK-C, and therefore were more complex presentations and so classified as mixed (see Figure 6.1).

Of participants identified as dysfunction syndrome (36%) following the initial McKenzie classification component of the MK-C, 78 tested positive for patho-anatomical sources of their symptoms with the MK-C, and were assigned an MK-C classification of FJS, SI, or mixed. The remaining 22% were negative for patho-anatomical sources of their symptoms and were assigned an MK-C classification of dysfunction syndrome (see Figure 6.1).

Following the MK-C, only 6% of participants remained inconclusive. Therefore, of participants identified as inconclusive (31%) following the initial McKenzie classification component of the MK-C, 38 were classified as FJS, SI, SIJS, or mixed on MK-C examination (see Figure 6.1).

6.3.3 Characteristics of the MK-C: Inter-classification comparison

Table 6.2 summarises the CLBP demographics and characteristics according to each MK-C classification. Participants classified as FJS were significantly older than those classified as discogenic ($p < .001$) or mixed ($p < .001$). Participants classified as discogenic had significantly higher RM ($p = .02$) and MSPQ ($p = .01$) scores than those classified as FJS. In contrast, there were no significant differences in RM and MSPQ scores between any other MK-C classification categories.

Table 6.2

Demographic Data and Inter-Classification Comparison of Chronic Low Back Pain Characteristics Between Mk-C Classifications

	Total N = 150	Discogenic N = 35	Dysfunction N = 12	Inconclusive N = 9	Facet joint syndrome N = 74	Spinal instability N = 6	Mixed N = 7
Female	97 (65%)	23 (66%)	7 (58%)	6 (67%)	48 (65%)	3 (50%)	6 (86%)
Male	53 (35%)	12 (34%)	5 (42%)	3 (33%)	26 (35%)	3 (50%)	1 (14%)
Symptom duration (months)	36 (52)	36 (54)	35 (94)	60 (141)	36 (47)	75 (171.5)	24(84)
Age ($\overline{x} \pm SD$)	56.7 ± 14.32	51.06 ± 14.43**	55.92 ± 13.95	57 ± 10.91	62.91 ± 12.48§**	47.83 ± 12.29	41.71 ± 8.42§
Body mass index (kg/m²) ($\overline{x} \pm SD$)	31.5 ± 7.27	29.84 ± 7.35	34.95 ± 7.52	29.52 ± 6.42	32.05 ± 7.17	28.65 ± 5.17	31.49 ± 6.38
ODI ($\overline{x} \pm SD$)	45.2 ± 14.54	49.76 ± 12.33	39.17 ± 13.85	46.81 ± 19.13	43.24 ± 14.38	53.63 ± 12.47	47.46 ± 9.4
RM△ *Mdn (IQR)*	54.1 (35.42)	66.67 (29.17)*	52.09 (31.28)	50 (41.67)	50 (34.35)*	64.58 (51.05)	54.17(29.17)
MSPQ *Mdn (IQR)*	20.5 (20.5)	30.8 (30.8)*	15.4 (21.8)	20.5 (37.2)	17.9 (12.8)*	37.15 (21.85)	30.08(28.2)
VAS *Mdn (IQR)*	44 (34.5)	59 (40)	42.5 (43.5)	45 (52.5)	41.5 (27)	67 (40.5)	35(61)

Note. Classifications with <5 subjects not reported. ODI = Oswestry Disability Index; RM = Roland–Morris disability questionnaire; MSPQ = Modified Somatic Perceptions Questionnaire; VAS = Visual Analogue Scale.

§Significant at p < 0.05; ** Significant at p < 0.001.†

McKenzie classification	MK-C examination criteria	MK-C classification
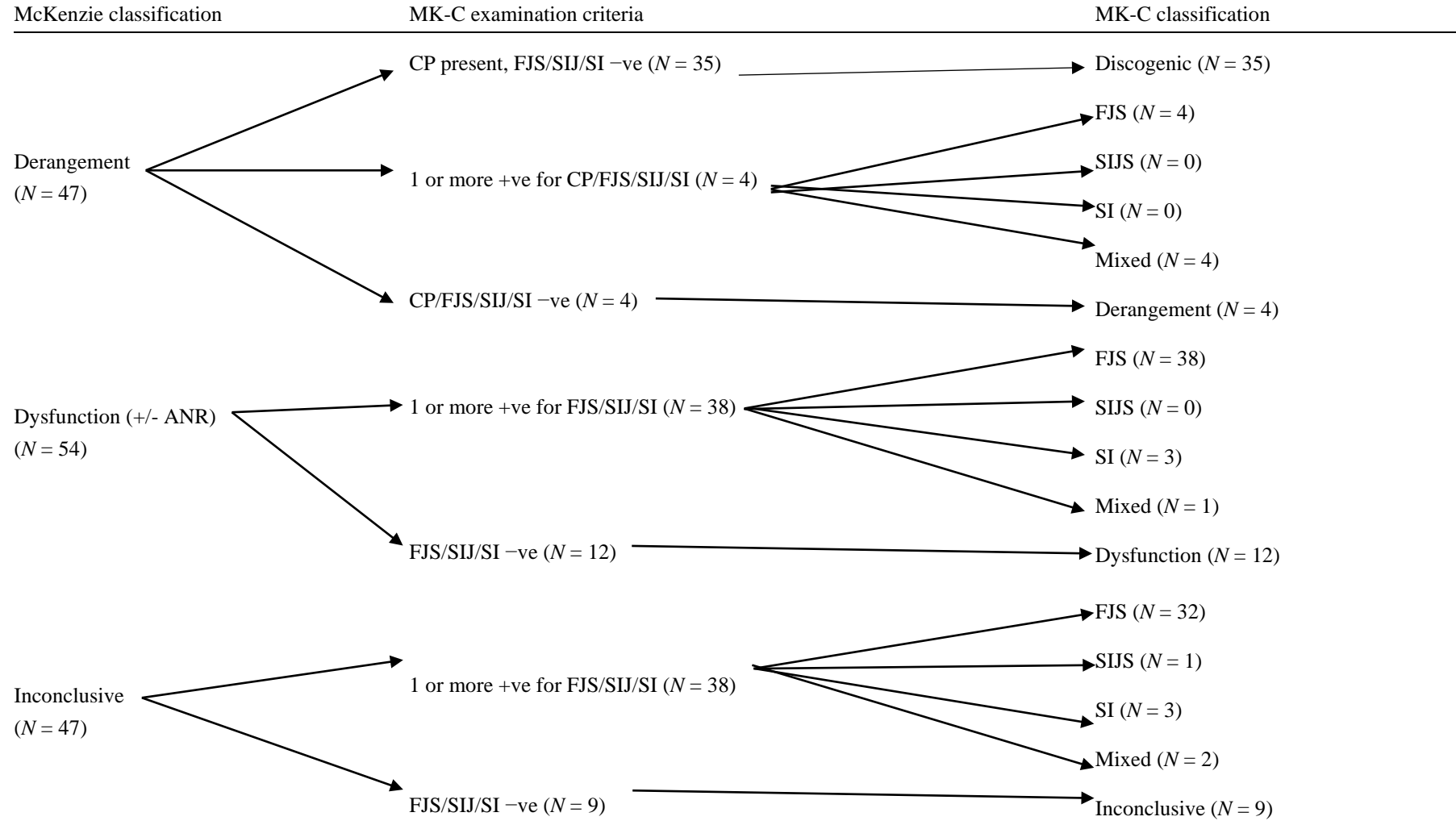
Figure 6.1. MK-C classification derived following the initial McKenzie assessment.

Similarly, there were no significant differences in ODI ($p = .09$), BMI ($p = .24$), VAS score ($p = .63$) or symptom duration ($p = .52$) between any MK-C classification categories.

6.4 Discussion

This study indicated that, using the MK-C, the most common classification in a CLBP population was FJS, with the least common SIJS, postural syndrome and SI. Consistent with previous studies that reported 0–2% postural pain (Clare et al., 2005; Eirikstoft & Kongsted, 2014; Hefford, 2008; Razmjou et al., 2000), this study found that less than 2% of participants classified as CLBP of postural origin.

Findings from this study identified that successful classification using the MK-C in a CLBP population can be achieved in 94% of cases at first attendance. Several consecutive attendances have been advocated to fully classify patients when exclusively applying a McKenzie Institute Lumbar Spine Assessment (McKenzie & May, 2003). Indeed, in this study the initial McKenzie Institute Lumbar Spine Assessment component of the MK-C resulted in an inconclusive classification of 31% of participants at first attendance. However, the MK-C appears to be more efficient as it provided a final classification at the first examination for the McKenzie syndromes of derangement without discogenic indicators, mechanically inconclusive and dysfunction. Therefore, the ability of the MK-C to classify the majority of participants at first examination was highlighted by this research and supports the use of this combined McKenzie assessment and patho-anatomical examination method to optimise classification of CLBP patients.

To the author's knowledge, the classification characteristics of a CLBP population examined using an McKenzie Institute Lumbar Spine Assessment or MK-C has not been described previously. Only two past studies have reported LBP

classification characteristics using a standardised classification system, but the systems used, the study populations (Eirikstoft & Kongsted, 2014; Hefford, 2008), and examiner experience (Eirikstoft & Kongsted, 2014) were dissimilar to the current study, limiting comparability of findings between this and those past studies.

The MK-C provided key findings, particularly in relation to participant age, disability and modified somatic pain perceptions among classification categories. Patients classified as FJS were significantly older than those classified as discogenic pain or a mixed presentation. These age-related classification differences between FJS and symptoms of intervertebral disc origin are explained and supported by previous evidence that degeneration of the disc precedes that of the facet joints (Butler, Trafimow, Andersson, & McNeill, 1990; Fujiwara et al., 1999; Miller, Schmatz, & Schultz, 1988).

The significantly higher RM and MSPQ scores for the discogenic classification compared with FJS is more difficult to explain, particularly as the difference in the mean RM scores for the FJS and discogenic classifications was 4/24, which, although statistically significant, may not be considered clinically significant (Roland & Fairbank, 2000). Additionally, on the basis of the centralisation of pain concept, higher modified somatic pain perception related to increased symptom duration has been documented (Ardic & Toraman, 2002). Conversely, no significant differences were reported in symptom duration between any classification categories in this cohort, suggesting that differences in MSPQ scores between discogenic and FJS were not attributable to symptom duration. Further, the mean MSPQ scores of all classification categories rated as 'distressed' (Main et al., 1992), indicating elevated somatic awareness in both classifications.

Consequently, further research is needed to explore the role of modified somatic pain perceptions in CLBP presentations.

The findings of this study highlight and support the utility of the MK-C for assessment and classification of CLBP. This enables physiotherapists to utilise diagnostic skills to classify CLBP patients at the first visit, while maintaining their unique ability to assess according to the fundamentals of movement, symptom provocation and function. However, certain MK-C categories had limited numbers which reduces the power of the conclusions which may be drawn and ultimately the statistical inference that may be taken from these group results. Also, the results are limited to the assessment of one physiotherapist, and the classification characteristics presented are specific to CLBP and the classification algorithm used for this study. Accordingly, results may not be reproduced by other physiotherapists, and cannot be generalised to other LBP subgroups or classification systems. Nevertheless, in the absence of previous evidence specific to this topic, the reported findings provide a unique contribution to knowledge on CLBP, and are relevant and clinically useful.

6.5 Conclusion

This study indicated that 94% of CLBP patients could be classified using the MK-C. The most common presentation in CLBP was FJS, with SIJS, postural syndrome and SI the least common. The MK-C examination could refine the McKenzie syndromes of derangement without discogenic indicators, mechanically inconclusive and dysfunction into a final classification based on patho-anatomical causes of patient symptoms. Age, RM and MSPQ were the only characteristics that varied among classification categories and appeared to be distinguishing characteristics of this population. Future studies should be conducted to establish inter-examiner reliability across several examiners.

Key points:

- Ninety-four per cent of CLBP patients could be classified using the MK-C on initial assessment.

- Some McKenzie syndromes could be refined into a final patho-anatomical classification using the MK-C examination algorithm.

- The dysfunction classification represented a greater proportion of the sample population, compared with previous studies in heterogeneous or acute LBP populations.

- Age, RM and MSPQ characteristics varied among classification categories and appeared to be distinguishing characteristics of CLBP.

6.6 What This Chapter Adds to Current Clinical Knowledge, and How It Informed Subsequent Stages of the Research

It was important to identify the classification characteristics of a CLBP population, as these had not been reported previously. Understanding the defining characteristics of clinical groups assists health practitioners to focus assessment, interventions and management specific to the group of interest. The study highlighted that some characteristics of CLBP patients are distinct from those of other LBP subgroups, not only demographically but also according to classification distributions. This unique contribution to knowledge will assist physiotherapists to profile patients during assessment, and focus treatment accordingly. In addition, the characteristics and classification profile reported provides new understanding, which can be used as a comparator for future studies.

The study sample reported in this study were all participants recruited to a more extensive study, reported in the succeeding chapter of this thesis (see Chapter 7). This chapter was therefore a prior analysis of the characteristics of the wider

study sample and its diagnostic classifications as they relate specifically to the MK-

C, and provided background and supported the succeeding study (see Chapter 7).

Chapter 7: Level of Agreement Between a Physiotherapist Using a Chronic Low Back Pain Examination Algorithm and Diagnostic Reference Standards

This chapter addresses the fourth research question (see Chapter 1): 'How well does the diagnostic classification of a physiotherapist using a comprehensive CLBP examination algorithm agree with diagnostic reference standards?'

The chapter includes content from 'Reply to the Letter to the Editor Regarding "Classification Characteristics of a Chronic Low Back Pain Population Using a Combined Mckenzie and Patho-Anatomical Assessment"' (Flavell, Gordon, & Marshman, 2017b; see Appendix 3).

This chapter reports the level of CLBP diagnostic classification agreement of two health professionals, a physiotherapist using the MK-C (detailed in Chapter 4) and available reference standards that comprised a CLBP specialist neurosurgeon's diagnosis derived from clinical examination and advanced medical imaging plus DSAI if clinically indicated.

7.1 Introduction

Chronic pain conditions are complex and require a multidisciplinary management approach (Kamper et al., 2015). Preferentially, CLBP is assessed and classified by a team of health professionals including but not exclusive to medical specialists, physiotherapists, nurses, social workers and psychologists. In such settings, shared terminology and communication is fundamental for optimal management (Cedraschi, Nordin, Nachemson, & Vischer, 1998). This highlights the importance of inter-professional diagnostic agreement in this complex group of patients to facilitate effective clinical reasoning, multidisciplinary management and

better patient outcomes (Cedraschi et al., 1998; Paskowski, Schneider, Stevans, Ventura, & Justice, 2011). Indeed, lower costs and good clinical outcomes were reported when a standardised LBP spinal care classification pathway was applied across multiple health providers including physiotherapists and neurosurgeons (Paskowski et al., 2011).

Although a purely patho-anatomical approach to CLBP management is generally unsupported (McCarthy et al., 2012), within a multidisciplinary environment, patho-anatomical-related terminology remains accepted and understood among health disciplines. This is in contrast to alternate physiotherapy-specific classification terminology, which is recognised and accepted within the profession but not universally understood by other health professionals. This has potential to hinder effective communication and understanding between medical specialists and physiotherapists.

Accordingly, the combined movement and patho-anatomical classification approach of the MK-C (Flavell et al., 2016) may be suitable for use by physiotherapists working in an inter-professional CLBP team environment. However, the level of diagnostic agreement between a physiotherapist using the MK-C and other health professionals within a multidisciplinary CLBP management team has not been evaluated.

Previously, a blinded validity study reported the diagnostic agreement and validity of two physiotherapists using a combined movement and patho-anatomical examination process against a reference standard of physician examination with access to diagnostic imaging and spinal injection (Laslett, McDonald, Tropp, Aprill, & Öberg, 2005). However, the definition of CLBP was not explicitly defined in the study. A mean duration of CLBP of 157.9 weeks was stated, but the wide standard

deviations reported made it unclear if participants' symptoms aligned with current CLBP definitions (Henschke et al., 2006). Additionally, the examining physician was an interventional radiologist, a health profession not routinely associated with a CLBP management team. Hitherto, there have been no reports of diagnostic agreement level between an experienced physiotherapist and a neurosurgeon specialised in CLBP and linked specifically with a secondary-care CLBP multidisciplinary management team. Neither have the classifications of the MK-C been compared with suitable CLBP diagnostic reference standards currently used in a CLBP secondary-care facility. Thus, a neurosurgeon clinical diagnosis (NCD) based on examination, with radiography and magnetic resonance imaging (MRI) findings, followed by a clinically indicated confirmatory spinal injection may be considered a suitable reference standard against which to determine diagnostic agreement. Particularly as evidence suggests there is no diagnostic gold standard for specific structural sources of CLBP (Bogduk, 2004).

Accordingly, the primary objectives of this secondary-care CLBP study were to report (1) diagnostic classification agreement between an experienced physiotherapist using the MK-C and a reference standard comprising an NCD, confirmed by diagnostic spinal injection when clinically indicated, and (2) the combined agreement of the physiotherapist using the MK-C and the NCD relative to the outcome of diagnostic spinal injections.

7.2 Method

Details of the setting, dates and ACTRN registration number of this prospective blinded cross-sectional study have been provided previously (see Chapter 6.2.1).

7.2.1 Ethics

Study approval was obtained from the Human Research Ethics Committees of James Cook University and Townsville Hospital (JCUH4387/HREC10QTHS53; see Appendix 4B). Prior to commencement of the study, participants received a detailed written and verbal explanation of the study and provided written consent.

7.2.2 Participants

A convenience sample of consecutive patients who attended the weekly specialist CLBP spinal injection clinic was recruited to the study. Patients were referred to the clinic by either a primary healthcare physician or a healthcare specialist.

7.2.3 Inclusion criteria

Participants were 18 years or older with CLBP, defined as symptoms between the 12th thoracic vertebra and the buttock crease, persistent for longer than 12 weeks (Reneman et al., 2006) and with or without lower limb symptoms (Jones et al., 2003), and had undergone MRI of the lumbar spine prior to attendance at the clinic.

7.2.4 Exclusion criteria

Exclusion criteria for both the physical examination and spinal injection were previous surgery of the lumbo-sacral spine; intolerance to the physical examination or local anaesthesia; previous allergy to contrast injection or iodine exposure; warfarin use during the week prior to the injection; coagulopathy with deranged clotting function and blood coagulation disturbances; corticoid incompatibility; pregnancy; inability to communicate; psychiatric disorders that might interfere with the examination and the participant's interpretation of injection effect; systemic illness or infection; history of spinal infection; abdominal surgery in the previous year; current litigation, insurance or other compensation claims; medical conditions

124

that result in CLBP, such as fibromyalgia and ankylosing spondylitis; presence of medical 'red flags' indicative of potentially serious medical conditions; and progressive neurological disturbance.

7.2.5 Baseline data and self-reported questionnaires

Participant age (yr), gender, weight (kg), height (m) and baseline symptom distribution using a body chart were recorded. An 11-point VAS (Huskisson, 1974) was completed to assess pain intensity. This measure has high sensitivity to detect change and good reliability in LBP populations (Hawker et al., 2011).

7.2.6 Examiner procedures

In a private room, an experienced physiotherapist examined and screened volunteers for inclusion and exclusion criteria, collected baseline and demographic data, examined participants using the MK-C, and assigned a diagnostic classification according to processes described previously in Chapter 4.4 (Flavell et al., 2016). For the purposes of this study, if in the presence of an FJS classification, neurological or neuro-dynamic testing produced or increased lower limb symptoms, a classification of facet joint syndrome (FJS) with associated sciatica/nerve root (NR) compromise/radicular symptoms was assigned. Additionally, because the MK-C classification of 'dysfunction' and 'derangement (centralisation phenomenon absent)' do not relate to a specific structural cause of CLBP, this classification was merged with the 'inconclusive' classification (see Table 7.1).

A consultant neurosurgeon trained in spinal injection procedure examined the participants, reviewed the MRI and, in some cases, conducted DSAI. An independently derived NCD was assigned following the clinical examination and review of radiographs (when available or required) and MRI. Following the NCD, the neurosurgeon conducted a diagnostic spinal anaesthetic injection (DSAI)

procedure, if clinically appropriate and with patient consent, to confirm or rule out

the NCD.

The overall reference standard against which the physiotherapy diagnostic

classification was compared was the NCD combined with the outcome of the

diagnostic spinal injection procedure. The physiotherapy examination, NCD and

spinal injection procedures were conducted sequentially within 60–90 min. No other

clinical interventions were conducted. The diagnostic classifications included in this

study and the associated reference standards are detailed in Table 7.1.

7.2.7 Blinding

The physiotherapist was blinded to all medical records, diagnostic imaging

and diagnostic injection information. Both examiners were blinded to the other's

clinical diagnosis. The examiners recorded their diagnoses on standardised forms.

The physiotherapist recorded the classification derived from the MK-C, and the

neurosurgeon the diagnosis derived from the NCD. The neurosurgeon recorded the

NCD prior to conducting DSAI. If participants CLBP had a mixed diagnostic

presentation, both examiners recorded their primary clinical diagnosis and injections

were conducted on the basis of the primary NCD.

For participants who received a spinal injection procedure, the neurosurgeon

also completed a separate form to record DSAI details including type, spinal level

and body side. All forms (physiotherapists MK-C classification, the NCD and DSAI

details) were placed in individual envelopes, sealed and stored securely by an

independent research assistant.

7.2.8 Injection procedures

All injections were conducted in a sterile procedure room (Plate 7.1).

Injection types included fluoroscopy guided facet joint injection (FGFJI), to confirm

Table 7.1

Diagnostic Classifications and Associated Reference Standards

MK-C diagnostic classification	Reference Standard
Discogenic +/- NR[*] compromise	Neurosurgeon clinical diagnosis + MRI[€]
FJS	Neurosurgeon clinical diagnosis with MRI plus FGFJI[≠]
FJS with NR[*] compromise	Neurosurgeon clinical diagnosis with MRI plus TFNRB[§]
SIJS	Neurosurgeon clinical diagnosis + MRI
SI	Neurosurgeon clinical diagnosis + x-ray + MRI
Postural syndrome/Muscular-ligamentous	Neurosurgeon clinical diagnosis + MRI
Structurally inconclusive (Derangement/Dysfunction/Inconclusive)	Neurosurgeon clinical diagnosis + MRI
Other joint pathology/chronic pain with no evidence of mechanical or musculoskeletal origin	Neurosurgeon clinical diagnosis + MRI

Note. FGFJI = intra-articular facet joint injection; FJS = facet joint syndrome; MRI = magnetic resonance imaging; NR = nerve root; TFNRB = trans-foraminal injection; SIJS = sacroiliac joint syndrome.

FJS, and trans-foraminal nerve root block (TFNRB) to confirm FJS with associated NR compromise. Participants were positioned in a prone position on a procedure bed. For FGFJI, the neurosurgeon instilled a 5 ml aliquot of a 20 ml solution consisting of 0.25% Marcaine mixed with 80 mg methylprednisolone. For TFNRB, 2 ml 0.25% bupivacaine mixed with 40 mg methylprednisolone was injected into the foramen. All injections were conducted under aseptic conditions, with fluoroscopy guidance under image intensifier control, which included the instillation of 1 ml iopromide non-ionic contrast in participants who were not allergic.

Using the spinal injection, the neurosurgeon applied a 'diagnostic differentiation' rationale to confirm or rule out potential patho-anatomical sources of CLBP. Injections were conducted with an 'intention to treat' to abolish or reduce the CLBP specific to a symptomatic spinal level and body side. Marcaine provided diagnostic information (i.e., if no pain reduction was recorded then that joint was excluded as a pain generator).

7.2.9 Outcome measurement

For participants who received a spinal injection, post-injection symptom response was recorded by the physiotherapist. Patients self-reported pain intensity on a VAS (Huskisson, 1974) within 30 min of the procedure. Participants were instructed to refrain from reporting pain from the injection site, and only rate their CLBP. The percentage change in the VAS score from baseline was calculated and recorded, and cases were categorised as 'Better' if a 75% or greater reduction in the baseline VAS score was achieved and 'Not Better' if the VAS score was reduced less than 75%. This cut-off has been reported previously (Laslett et al., 2004; Revel et al., 1998) and was chosen because low false-positive diagnostic rates have been associated with similarly strict reference standard pain reduction levels (Hancock et

al., 2007). The neurosurgeon was blinded to the outcome of the DSAI and the patient self-reported VAS score.

Plate 7.1. The Townsville Hospital sterile procedure room.

Accordingly, the following definitions confirmed a positive or negative diagnostic response to spinal injection:

- FGFJI or TFNRB and 'Better' => Diagnosis '+ve FJS' or '+ve FJS with NR compromise', respectively

- FGFJI or TFNRB and 'Not Better' => Diagnosis '−ve FJS' or '−ve FJS with NR compromise', respectively.

7.2.10 Un-blinding data to determine examiner agreement

At completion of data collection, blinded injection and diagnosis forms were revealed by the independent research assistant, and matched for each participant. A

predetermined and jointly developed matching protocol among the physiotherapist MK-C classification, NCD, and reference standard (NCD plus outcome of DSAI) was applied by the research assistant, to ensure unbiased and independent MK-C diagnosis, NCD and reference standard matching. Where response to spinal injection was 'Not Better', the NCD was not confirmed and therefore the reference standard diagnosis was 'inconclusive'. The examiner diagnoses were condensed to eight categories, which reflected the accepted diagnostic classification terminology used in the CLBP clinic (see Table 7.1).

7.2.11 Physiotherapist MK-C classification and reference standard outcome agreement

Diagnostic 'Agreement' or 'Disagreement' between the physiotherapist's MK-C diagnostic classification and the reference standard was recorded using the following definitions:

- MK-C classification and reference standard agree => 'Agree'

- MK-C classification and reference standard do not agree => 'Disagree'.

7.2.12 Physiotherapist MK-C classification and neurosurgeon clinical diagnosis combined agreement relative to the outcome of diagnostic spinal anaesthetic injection

Analysis of combined diagnostic agreement of the physiotherapist's MK-C classification and the NCD compared with the outcome of the DSAI was reported according to the following definitions:

- Examiners 'Agree' and '+ve FJS' or '+ve FJS with NR compromise' => 'Agree/Better' (MK-C true positive)

- Examiners 'Agree' and '−ve FJS' or '−ve FJS with NR compromise' => 'Agree/Not Better' (MK-C false positive)

130

- Examiners 'Disagree' and '+ve FJS' or '+ve FJS with NR compromise'

  => 'Disagree/Better' (MK-C false negative)

- Examiners 'Disagree' and '−ve FJS' or '−ve FJS with NR compromise'

  => 'Disagree/Not Better' (MK-C true negative).

7.2.13 Statistical analysis

A 70% diagnostic agreement among the physiotherapist MK-C classification, NCD, and reference standard was expected, resulting in an effect size index of 0.4. Therefore, based on α = .05 and to achieve a power of 80% or greater, the estimated minimum sample size was 50 participants.

All data were imported to SPSS statistical analysis software, version 22 (SPSS Inc., Chicago, IL). Descriptive statistics were generated and numeric variables were analysed for normality of distribution and reported using mean and SD ($\overline{x} \pm SD$) or Mdn with IQR. Categorical variables for physiotherapist MK-C classification and reference standard agreement, and physiotherapist MK-C classification and NCD agreement relative to the outcome of spinal injection, were generated and reported as proportions (%). All data were complete and no indeterminate MK-C, NCD, or reference standard outcomes existed.

A diagnostic classification percentage of observed agreement (%) and chance corrected agreement between the MK-C classification and the reference standard using a kappa coefficient was calculated and reported with 95% CI. Kappa values were interpreted as poor (<.0), slight (.00–.2), fair (.21–.4), moderate (.41–6), substantial (.61–.8) and almost perfect (.81–1; Landis & Koch, 1977). For those participants who received DSAI ($N = 50$), chi-square cross-tabulation was conducted to report the combined agreement of the physiotherapist MK-C classification and NCD relative to the outcome of DSAI.

7.3 Results

The recruitment process and flow of participants through this study is detailed in Figure 7.2. Between July 2012 and March 2014, 197 patients who attended the CLBP clinic volunteered to participate in the study. Inclusion and exclusion criteria were applied by the physiotherapist, with 47 excluded because of previous surgery of the lumbo-sacral spine ($N = 9$); current litigation, insurance or other compensation claims ($N = 4$); non-lumbar or pelvic pain ($N = 6$); ankylosing spondylitis or systemic lupus erythematosus ($N = 2$); and inability to tolerate the physical examination because of poor standing balance or other limiting co-morbidities ($N = 26$).

All eligible participants were examined by both examiners ($N = 150$). However, unavoidable changes to clinic schedules and unforeseen medical priorities resulted in the neurosurgeon omitting to complete a diagnosis form in some cases ($N = 51$). Additionally, if an updated MRI was required ($N = 1$) or patient symptoms were attributable to psychosocial or biological factors unrelated to LBP ($N = 6$), no diagnosis was recorded by the neurosurgeon. These cases were subsequently excluded, and the remaining participants were included in the physiotherapist MK-C classification and reference standard agreement analysis ($N = 92$). Demographic and CLBP characteristics of the study population are presented in Table 7.2. There were no adverse events during any physiotherapy or neurosurgeon examination. There were no significant age, body mass index (BMI) or baseline VAS score differences among participants where the reference standard and physiotherapist agreed or disagreed ($p > .05$).
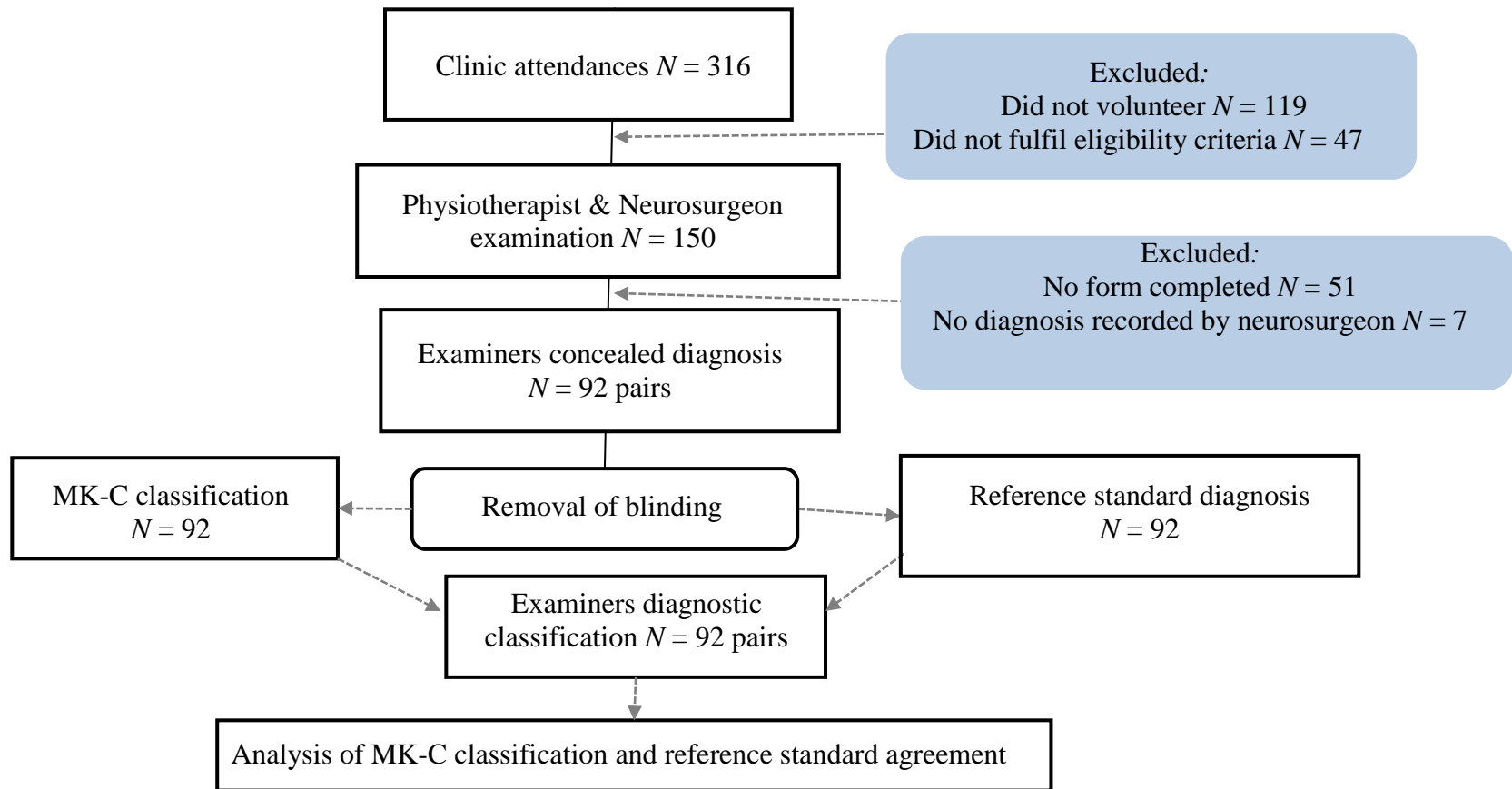
Figure 7.2. Flow chart of study recruitment, blinding and matching process.

7.3.1 MK-C classification and reference standard agreement

The physiotherapist classification (%) and reference standard diagnoses (%) are shown in Table 7.3. Overall diagnostic classification agreement between the physiotherapist and the reference standard was 51% ($k = .22$, CI [0.08, 0.42]), which indicated 'Fair' agreement.

7.3.2 Outcome of diagnostic spinal anaesthetic injections

Diagnostic injections to confirm FJS or FJS with NR compromise were conducted in 55% of cases. FGFJI was conducted in 48% ($N = 44$) and TFNRB in 7% ($N = 6$). Of participants who received a spinal injection ($N = 50$), 70% were 'Better' and 30% 'Not Better'. Those who were 'Not Better' had a significantly higher pre-injection pain intensity, compared with those who were 'Better' ($p = .001$, CI [−29.39, −8.12]). No other significant differences in pre-injection characteristics existed.

7.3.3 Physiotherapist MK-C classification and neurosurgeon clinical diagnosis combined agreement relative to the outcome of diagnostic spinal anaesthetic injection

The cross-tabulation of combined MK-C and NCD relative to DSAI outcome is provided in Table 7.4. The NCD diagnosis was confirmed by the reference standard spinal injection ('+ve FJS' or '+ve FJS with NR compromise') when there was MK-C and NCD diagnostic agreement in 46% of cases, indicating that jointly the examiners correctly diagnosed the source of CLBP (MK-C true +ve). However, the NCD diagnosis was not confirmed by the reference standard spinal injection ('−ve FJS' or '−ve FJS with NR compromise') when there was MK-C and NCD diagnostic agreement in 18% of cases, indicating that jointly the examiners incorrectly diagnosed the source of CLBP (MK-C false +ve).

Table 7.2

Demographic and Clinical Characteristics

	All participants	Examiners agree	Examiners disagree	*p* value (CI)
Total	92 (100%)	47 (51%)	45 (49%)	
Female	66 (72%)	32 (68%)	34 (76%)	
Male	26 (28%)	15 (32%)	11 (24%)	
Age (yr) ($\bar{x} \pm SD$)	56.25 ± 15.36	58.17 ± 15.17	54.24 ± 15.47	.22 [−10.27, 2.42]
Body mass index (wgt/hgt^2)$^\#$ ($\bar{x} \pm SD$)	31.41 ± 7.30	31.22 ± 8	31.59 ± 6.57	.81 [−2.68, 3.44]
Baseline VAS ($\bar{x} \pm SD$)	49.25 ± 20.97	46.4 ± 21.91	52.22 ± 19.74	.19 [−2.84, 14.47]
Duration of symptoms (months) *Mdn* (*IQR*)	36 (53)	36 (50)	30 (61.5)	.86 [−15, 12]

Note. CI = confidence interval; VAS = visual analogue scale.

The NCD was confirmed by the reference standard spinal injection ('+ve FJS' or '+ve FJS with NR compromise'), and the MK-C and NCD disagreed, in 24% of cases, indicating that the NCD was correct but that the physiotherapist's MK-C classification was incorrect (MK-C false −ve). In these cases of examiner disagreement, where the neurosurgeon diagnosed intra-articular FJS correctly ($N = 9$), the physiotherapist's MK-C classification was inconclusive ($N = 4$), discogenic ($N = 4$) or SI ($N = 1$). Additionally, where the neurosurgeon diagnosed FJS with NR compromise correctly ($N = 3$), the physiotherapist's MK-C classification was discogenic ($N = 2$) or intra-articular FJS in the absence of NR compromise ($N = 1$).

In 12% of cases, the NCD diagnosis was not confirmed by the reference standard spinal injection ('−ve FJS' or '−ve FJS with NR compromise') when the MK-C and NCD disagreed, indicating the NCD to be incorrect. In five of these cases,

Table 7.3

Diagnostic Classifications (*N* = 92)

	Physiotherapist	Reference standard
Discogenic +/- nerve root compromise	20 (22%)	16 (17%)
Facet joint syndrome	53 (58%)	51 (55%)
Facet joint syndrome with nerve root compromise	1 (1%)	6 (7%)
Sacroiliac joint syndrome	1 (1%)	1 (1%)
Spinal instability	7 (8%)	1 (1%)
Postural syndrome/Muscular-ligamentous	1 (1%)	1 (1%)
Structurally inconclusive (Derangement/Dysfunction/Inconclusive)	9 (10%)	15 (16%)
Other joint pathology/Chronic pain with no evidence of mechanical or musculoskeletal origin	0 (0%)	1 (1%)

Table 7.4

Cross-Tabulation Combined Diagnoses Relative to Spinal Injection Outcome

		Spinal injection outcome		
		Better	Not Better	Total
MK-C & NCD Agreement	Agree	23 (46%)	9 (18%)	32 (64%)
	Disagree	12 (24%)	6 (12%)	18 (36%)
Total		35 (70%)	15 (30%)	50 (100%)

Note. NCD = neurosurgeon clinical diagnosis.

the neurosurgeon incorrectly diagnosed intra-articular FJS, but the physiotherapist classified SI ($N = 2$), and discogenic ($N = 3$). In the remaining case, the neurosurgeon incorrectly diagnosed FJS with NR compromise, but the physiotherapist classified the patient as intra-articular FJS (no NR compromise). Although the physiotherapist using the MK-C correctly ruled out the source of the CLBP (MK-C true −ve), in these cases, the physiotherapist's MK-C classification could not be confirmed.

7.4 Discussion

This blinded study reported 'Fair' agreement between the diagnostic classification of an experienced physiotherapist using the MK-C examination algorithm and a reference standard of diagnosis by a neurosurgeon derived from clinical examination, MRI and, when clinically indicated, the outcome of diagnostic spinal injections. Further, this blinded study found that, in participants who received a diagnostic FGFJI or TFNRB, examiners joint diagnostic agreement of FJS or FJS with NR compromise was confirmed in 46% of cases.

7.4.1 MK-C classification and reference standard agreement

Currently, little evidence exists to support the validity of any complete physiotherapy examination process applied to CLBP patients managed in a

multidisciplinary setting. The results of this study indicated 'Fair' agreement between a physiotherapist using a comprehensive CLBP examination algorithm to provide a diagnostic classification and available reference standards.

Although these levels of agreement are not optimal, they provide an understanding of the diagnostic value of the blinded clinical examination process used in this study when compared with the diagnosis achieved with clinical examination, imaging results and diagnostic injection procedures. Despite the differences between this study and that of Laslett, McDonald, et al. (2005) outlined earlier in this article, both studies reported a 'Fair' level of after-chance diagnostic agreement between a physiotherapy clinical examination and a reference standard.

The MK-C examination process used in this study reflects physiotherapy practice where classification founded on a clinical examination predominates. The sample population in this study was representative of CLBP patients seen by physiotherapists in secondary-care clinical practice, and is reflective of CLBP populations studied previously (Ferreira et al., 2013; Iversen et al., 2013; Viniol et al., 2013).

7.4.2 Physiotherapist MK-C classification and neurosurgeon clinical diagnosis combined agreement relative to the outcome of diagnostic spinal injection

Outcome of combined examiner diagnoses compared with the response to FGFJI and TFNRB provided information on the diagnostic value of a team-based examination approach, which is typical in multidisciplinary management of CLBP, yet has not previously been reported in this way. Compared with FGFJI and TFNRB as a diagnostic reference standard, the examiners joint blinded diagnosis was correct in almost half of cases. Eighteen per cent of cases that did not actually have FJS or FJS with NR compromise, were falsely diagnosed by both examiners. Despite the

138

neurosurgeon having access to MRI results, this rate of false-positive diagnoses is unsurprising, as high rates of false-positive diagnoses attributed to MRI in LBP have been well documented (Boden, Davis, Dina, Patronas, & Wiesel, 1990; Hancock et al., 2012; Jensen et al., 1994; Tawa, Rhoda, & Diener, 2016; Wassenaar et al., 2012).

While the physiotherapist incorrectly ruled out FJS or FJS with NR compromise in almost one-quarter of cases when the condition was actually confirmed present, the neurosurgeon provided the correct diagnosis. These results may indicate that, although not perfect as a diagnostic tool, access to MRI results combined with clinical examination may have optimised the NCD. Nevertheless, the physiotherapist correctly ruled out FJS or FJS with NR compromise in 12% of the injected cases.

Results indicated that, when examiners disagreed, no consistent pattern of contrasting classification by the physiotherapist using the MK-C was apparent. However, when disagreement occurred, the physiotherapist appeared less likely to correctly rule out FJS or FJS with NR compromise using the MK-C. The clinical implications of this are that patients may be classified with FJS or FJS with NR compromise incorrectly using the MK-C. However, these results appear to contradict findings reported previously using clinical predictors for FJS, which were included in the MK-C algorithm, with similarly high post-injection VAS score cut-off rates. Specifically, Laslett, McDonald, et al. (2006), in a study of 151 CLBP patients, reported that a cluster of clinical tests for FJS had most predictive value to rule out a positive response to FGFJI. Therefore, clinicians should be mindful of this contrasting evidence when applying these findings in clinical practice.

The author also acknowledges that, in this study, the examination processes had a biological focus in a condition inherently biopsychosocial in nature. This

potentially reduces the diagnostic accuracy of both examiners. Given these clinical examination challenges for physiotherapists and medical specialists working with CLBP patients, the 'Fair' examiner agreement and proportion of joint examiner true positive and negative diagnoses reported in this study are unsurprising. Indeed, they substantiate suggestions from another CLBP study with similar demographic characteristics that clinicians may need to apply a variety of examination systems to address psychological in addition to biomedical facets of LBP (McCarthy et al., 2012).

Although this may be theoretically fitting, rarely are individual clinicians skilled in all of the biological, psychological and social aspects of patient examination. Customarily, clinicians conduct evidence-based examination specific to their clinical expertise and scope of practice, aiming to conduct an examination that is the best available. The inclusion of a McKenzie classification (MII, 2005) and patho-anatomical test clusters previously validated against reference standards (Iversen et al., 2013; Laslett, 2008; Laslett, Aprill, et al., 2006; Laslett, Aprill, et al., 2005; Laslett, McDonald, et al., 2006; Laslett et al., 2004; Laslett, Öberg, et al., 2005; Laslett et al., 2003) in the composite MK-C algorithm used in this study showed 'Fair' agreement with reference standard diagnosis. Considering the multifaceted complexity of CLBP, the ability of the MK-C to classify patients according to functional, movement-based and patho-anatomical deficit appears clinically useful.

7.4.3 Limitations of this study

Despite providing new knowledge, this study has limitations. The sample population was specific to CLBP participants treated within a secondary-care setting in a multidisciplinary pain clinic; therefore, findings may not be related to acute or

subacute LBP, or patients seen in other healthcare settings. This was a single-centre study; hence, results only represent the demographic of CLBP referred to the medical facility described in this thesis. The examinations were conducted by the same physiotherapist and neurosurgeon, and so results may not represent the classification or diagnoses of all physiotherapists or neurosurgeons.

This study was conducted using reference standards available at the study site. For CLBP of discogenic origin, provocation discography is the preferred diagnostic method (Stout, 2010), and would have provided a more suitable reference standard than MRI. However, it was unavailable in this study. It is acknowledged that MRI, FGFJI and TFNRB are not diagnostic gold standards for CLBP. A study of CLBP patients treated with FGFJI identified that the diagnostic false-positive rate was 25–44% (Boswell et al., 2015). Notwithstanding, with the exception of provocative discography for discogenic pain, MRI and diagnostic spinal injection are accepted clinical diagnostic reference standards for other sources of spinal symptoms (Curatolo & Bogduk, 2010; Ngan et al., 2016; Seising, 2006). Thus, MRI, FGFJI and TFNRB were the optimal diagnostic methods available for this study, with a 75% or greater post-injection VAS score reduction cut-off applied to confirm FJS or FJS with NR compromise. This cut-off was considered suitable to avoid the high false-positive diagnostic rates reported previously with lower cut-offs levels (Hancock et al., 2007).

Fifty-one participants did not receive a NCD leading to unavoidable bias. Nevertheless the sample size achieved was 92 participants. Of the 92 participants in this study, only cases with a clinical indication for spinal injection and an 'intention to treat' received diagnostic spinal injections, according to the NCD. However, a further seven participants were classified with FJS or FJS with NR compromise by

the physiotherapist only. Because the examiners were blinded to each other's diagnoses, no FGFJI or TFNRB was conducted in these cases, and so it was not possible to confirm or refute the physiotherapists MK-C classification of FJS or FJS with NR compromise in those seven cases. Accordingly, the author acknowledges that conclusions related to any specific diagnostic value of the physiotherapy examination within the context of the physiotherapist MK-C classification and NCD combined agreement relative to the outcome of DSAI are subject to unavoidable verification bias.

7.4.4 Recommendations for future research

Further research using the MK-C algorithm across multiple sites and examiners is required to provide further knowledge of its efficacy as a classification process in CLBP. Refinements to the process should be conducted to improve inter-examiner agreement in CLBP. Studies conducted in other groups of less complex LBP patients may demonstrate greater levels of diagnostic agreement. In addition, diagnostic accuracy should be established in the absence of verification bias.

7.5 Conclusion

Improving inter-professional diagnostic agreement in CLBP should be a clinical priority to facilitate optimal management. Diagnostic agreement on the source of CLBP was greater than chance and rated 'Fair' in this study of patients referred to a secondary-care clinic. Jointly, an NCD and physiotherapy clinical examination using the MK-C provided 46% true-positive diagnoses for FJS or FJS with NR compromise in participants who received DSAI. Further research to refine the MK-C and evaluate its efficacy as an examination process for CLBP is recommended.

Key points:

- Diagnostic agreement between a physiotherapist using the MK-C and available reference standards rated 'Fair'.

- Jointly, the examiners correctly diagnosed FJS or FJS with NR compromise in 46% of cases that received FGFJI or TFNRB.

- Jointly, the examiners incorrectly diagnosed FJS or FJS with NR compromise in 18% of cases that received FGFJI or TFNRB.

7.6 What This Chapter Adds to Current Clinical Knowledge, and How It Informed Subsequent Stages of the Research

It was important to report the level of classification agreement when the MK-C algorithm was applied in a multidisciplinary CLBP setting to facilitate effective clinical reasoning and target patient management appropriately. The MK-C was developed in consideration of current best evidence on physiotherapy classification in CLBP, combined with the need to provide an examination process that classified in language that traversed health disciplines within the CLBP management team.

This study focused on the physiotherapy and neurosurgical diagnostic relationship. Understanding the level of agreement between the physiotherapist and the neurosurgeon within the CLBP clinic setting provided insight into how well the diagnosis of the two disciplines converged. This is important because, in a team environment, a medical decision to conduct a spinal injection with 'intention to treat' is often reinforced by the findings of physiotherapy clinical examination or a patient's response to preceding physiotherapy intervention. Further, the injecting doctor will often advocate physiotherapy to facilitate spinal rehabilitation following the procedure. It was therefore important to identify if the MK-C classification by the physiotherapist and a treating neurosurgeon's diagnosis had acceptable levels of

agreement. Future studies need to establish improved levels of diagnostic accuracy, but this was outside the scope of this dissertation.

Pre- and post-injection interaction between physiotherapists and injecting medical specialists is key in CLBP management. Despite the complexity of CLBP, the knowledge gained from this study suggested that methods of assessment that demonstrate a shared diagnostic language between health disciplines can be achieved. The ensuing studies of this research project investigated the post-injection outcomes related to function of spinal stabilising muscles in CLBP, which as yet has been subject to limited research relevant to physiotherapy.

Chapter 8: Measurement of the Transversus Abdominis in Chronic Low Back Pain Using Real-Time Ultrasound

To report the post-injection outcomes related to function of spinal stabilising muscles in CLBP using real-time ultrasound imaging (RTUI), the potential methodological confounders of optimal transversus abdominis (TrA) measurement required investigation.

This chapter addresses the fifth research question (see Chapter 1): 'In a CLBP study sample, what is the utility of a novel equipment developed to address TrA RTUI acquisition confounders of uncontrolled probe force, inclination and roll, and is TrA RTUI measurement of a single physiotherapist using the novel equipment reliable?'

The study reported in this chapter continued the stage based approach of this thesis, and was required to justify the methodology applied in the final study of this project. This study reviews, reports, and mitigates potential methodological confounders of uncontrolled 'free hand' RTUI to measure TrA thickness. It presents new methodology to address some of these confounders. The methods presented utilise a new 'force probe device' technology, which is capable of standardising RTUI probe parameters of force, inclination and roll across repeated imaging sessions.

This chapter reports a study entitled 'Measurement of Transversus Abdominis Activation in Chronic Low Back Pain Patients Using a Novel Standardised Real-Time Ultrasound Imaging Method' (Flavell, Gordon, & Marshman, 2017a), submitted for publication.

8.1 Background to Methodological Issues in Real-Time Ultrasound Imaging Transversus Abdominis Measurement in Chronic Low Back Pain

Over 30 years ago, the use of RTUI as an objective tool to measure abdominal muscle dimensions, to aid rehabilitation and research in LBP, was reported (Krag et al., 1987). Since then, the measurement of one specific abdominal muscle, the TrA, has become preeminent because of its key role in spinal stabilisation (Panjabi, 1992; Tesh et al., 1987), and because of the relationship of TrA dysfunction with LBP (Critchley & Coutts, 2002; Ferreira, Ferreira, & Hodges, 2004; Hides et al., 2009; Kiesel et al., 2008).

To date, there has been a consensus opinion that the TrA demonstrates thickness and feed-forward activation deficit during asymptomatic periods between recurrent episodes of LBP (Ferreira et al., 2004), and when LBP persists (Critchley & Coutts, 2002; Ferreira et al., 2004; Hides et al., 2009; Kiesel et al., 2008). Therefore, measurement and evaluation of TrA-C deficit in CLBP has clinical value.

Because of its validity, safety, relatively low cost and non-invasiveness (Koppenhaver, Hebert, Parent, & Fritz, 2009), the use of RTUI to measure TrA dysfunction has significantly increased over time. This has resulted in substantial research comparing differences in TrA thickness and transversus abdominis activation (TrA-C) within (Critchley & Coutts, 2002; Hides, Wong, Wilson, Belavý, & Richardson, 2007; Roddey, Brizzolara, & Cook, 2008; Watanabe, Kobara, Yoshimura, Osaka, & Ishida, 2014) and between (Ferreira et al., 2004; Hides et al., 2009; Kim, Cho, Goo, & Baek, 2013; Rostami et al., 2015) symptomatic and asymptomatic individuals. RTUI has demonstrated TrA-C anomalies and TrA morphological changes in patients with both acute and chronic LBP (Rostami et al.,

146

2015; Whittaker& Stokes, 2011), and has been used to demonstrate the efficacy of some therapeutic interventions to improve TrA function (Raney, Teyhen, & Childs, 2007; Streicher, Mätzold, Hamilton, & Wagner, 2014). However, the role of TrA to predict the course or recurrence of LBP is still the subject of debate (Wong, Parent, Funabashi, Stanton, & Kawchuk, 2013).

Although TrA thickness, length and volume can be measured using RTUI, TrA thickness at rest (RTrA), relative to that when fully contracted (CTrA), is primarily used to calculate the principle parameter of TrA-C (Koppenhaver, Parent, et al., 2009; Lariviere et al., 2013). RTUI TrA-C measurement has however been characterised by widely variable and generally sub-optimal examiner reliability (Costa, Maher, Latimer, & Smeets, 2009). Specifically, intra-class correlation coefficients (*ICC*s) for RTrA and CTrA have ranged from .41 to .97, while *ICC*s for TrA-C have ranged from .34 to .88 (Costa, Maher, Latimer, Hodges, & Shirley, 2009; Gnat et al., 2012; Hides, Miokovic, Belavý, Stanton, & Richardson, 2007; Lariviere et al., 2013; Mannion et al., 2008; Teyhen et al., 2005).

Despite evidence of sub-optimal reliability for RTUI TrA measurement, a systematic review of literature to evaluate methodological consistency and reported reliability for RTUI TrA measurement in CLBP had not been conducted. Therefore, the primary investigator of this PhD project supervised a student physiotherapist's honour's research to conduct the systematic review. This study by Whittle et al. (2017) identified that only two prior studies had reported TrA-C measurement reliability in CLBP (Costa, Maher, Latimer, Hodges, & Shirley, 2009; Mannion et al., 2008). *ICC*s reported in these studies ranged from .32 to .72.

Such variability is considered solely dependent upon the failed application of rigorous application protocols (Perkin, Bond, Thompson, Woods, & Smith, 2003),

yet a consensus regarding such protocols is currently lacking, and is problematic

given that RTUI is a 'free-hand' procedure (Costa, Maher, Latimer, & Smeets, 2009;

Hebert, Koppenhaver, Parent, & Fritz, 2009; Whittle et al., 2017).

Specific to CLBP, Whittle et al. (2017) found that sub-optimal

methodological consistency and measurement reliability for RTUI TrA in CLBP

prevails. To date, research has focused on healthy populations, and acute or subacute

LBP, with less focus on CLBP. Further, Whittle et al. (2017) suggested that the level

of reported reliability and methods currently used for RTUI TrA measurement in

CLBP limit the clinical and research application in this group of patients, and that

there is a need for researchers to improve methodological standardisation for CLBP

studies.

8.2 Identified Confounders of Real-Time Ultrasound Imaging Transversus Abdominis Measurement in Chronic Low Back Pain

Guided by the findings of Whittle et al. (2017), two TrA measurement

confounders were apparent, and predominate in CLBP. The first relates to physical

limitations associated with the CLBP population. While many RTUI TrA reliability

studies have focused on healthy populations (Costa, Maher, Latimer, & Smeets,

2009), patients with CLBP typically have increased body mass index (BMI)

(Cimolin et al., 2011; Heuch et al., 2010), and the associated increased adiposity

potentially thwarts the acquisition of clear TrA images (Brahee et al., 2013; Ortiz et

al., 2012). Increased abdominal wall compliance, and difficulties in palpating

reference anatomical landmarks, also present RTUI challenges specific to CLBP

(Thanh Le et al., 2015).

The second factor relates to inconsistences in probe application and

orientation, inherent with RTUI. Uncontrolled probe force, inclination and roll are

currently a feature of 'free-hand' scanning methods. Probe force has been subjectively classified in free-hand RTUI studies as 'gentle', 'light' or 'minimal' (Costa, Maher, Latimer, Hodges, & Shirley, 2009; Ota & Kaneoka, 2011; Unsgaard-Tøndel, Lund Nilsen, Magnussen, & Vasseljen, 2012), but this is not a quantitative evaluation and is sub-optimal. Maintenance of probe inclination perpendicular to the TrA during RTUI is a prerequisite to identify the hyper-echoic TrA fascial borders between which TrA thickness is measured (Perkin et al., 2003). Moreover, variable probe-to-skin force may distort TrA dimensions (Ishida & Watanabe, 2012; Perkin et al., 2003). Two free-hand RTUI studies have attempted to standardise probe orientation and skin force by securing the probe within a foam housing or fixing the probe to the participant with a belt (Lariviere et al., 2013; Mannion et al., 2008); TrA-C *ICC*s reported in these studies were poor to moderate (.32–.62; Lariviere et al., 2013; Mannion et al., 2008). Therefore, there was a critical need to conduct further research to establish RTUI TrA method standardisation aimed to improve measurement reliability for future CLBP studies.

8.3 Methods to Standardise Real-Time Ultrasound Imaging of the Transversus Abdominis

It was important to demonstrate high levels of test–retest reliability of RTUI TrA measurement for the subsequent pre–post intervention study of this research (see Chapter 9). Technology that standardised, quantified and subsequently better controlled probe force, inclination and roll was required. However, this is not a feature of any existing RTUI machine. Therefore, it was important to explore the availability of technologies with these standardising capabilities. If such technology could be combined with an RTUI machine to provide force inclination and roll data in real-time during scanning, then these parameters could be reproduced exactly

across repeated imaging sessions and between examiners, thereby standardising images for repeated measurement purposes.

In 2014, investigation revealed that a new technology known as a 'force probe device' (FPD) had been developed (Gilbertson & Anthony, 2013) by a bioengineering PhD student. Its principle aim was to show vascular tissue deformation, and subsequent measurement confounding associated with fluctuations in RTUI probe force and orientation during ultrasound scanning of the brachial artery (Gilbertson & Anthony, 2013). In principle, this FPD technology could standardise probe force, inclination and roll during RTUI by permitting continuous visual feedback to the examiner.

Technical details of the FPD have been reported elsewhere (Gilbertson & Anthony, 2013). This technology had potential to optimise TrA measurement reliability, so the bioengineering researchers who developed the FPD were contacted, and further discussion resulted in a signed research collaboration agreement between James Cook University and Massachusetts Institute of Technology (MIT OSP 24474 J Cook U non-profit SRA 20141219). A subsequent Research Infrastructure Block Grant was successful to fund purchase of the FPD (see Appendix 5).

8.4 Study Objectives

It was hypothesised that the FPD technology had potential to offset probe application and orientation confounders in this CLBP study population, and was required to improve the clinical validity and reliability of RTUI TrA measurement for this project and future research.

The objectives of this study were (1) to investigate the utility of using a FPD to standardise force, inclination and roll of the ultrasound probe during RTUI in a CLBP population, (2) to report this method for image acquisition and interpretation

and (3) to calculate the intra-examiner reliability for the standardised method prior to its application in the final study of this research.

8.5 Method

8.5.1 Study design

This was a blinded intra-examiner reliability study conducted in a multicultural specialist back pain clinic in a developed western country, with secondary day procedure facilities, as part of an existing registered prospective study (ACTRN: 12613000267752). Participants were referred to the clinic by either a primary healthcare physician or a healthcare specialist. Potential participants were consecutively approached as they attended the clinic between July and August 2015. Study approval was obtained from the Human Research Ethics Committees of James Cook University (H4387), Townsville Hospital (HREC10QTHS53; see Appendix 4B) and Mater Health Services (MHS20150512-07; see Appendix 4C). An explanation and information sheet detailing aims and methods of the study was provided to participants, and written consent was gained prior to study commencement. The primary investigator remained nearby and accessible throughout.

8.5.2 Inclusion and exclusion criteria

Participants were 18 years of age or older and presented with CLBP that adversely influenced their functional status. CLBP was defined as pain between the level of the 12th thoracic vertebra and the buttock crease, with or without associated lower limb symptoms (Jones et al., 2003), and present for longer than 12 weeks (Reneman et al., 2006).

Participants were excluded if they presented with any of the following criteria: previous surgery of the lumbo-sacral spine; pregnancy; inability to

communicate; psychiatric disorders that might interfere with the participant's interpretation of instructions; systemic illness or infection; tumour; trauma; fracture; abdominal surgery in the previous year; current litigation, insurance or other compensation claims; medical conditions that resulted in CLBP, such as fibromyalgia and osteoporosis; presence of medical 'red flags' indicative of potentially serious medical conditions; and progressive neurological disturbance.

8.5.3 Examiner

The primary investigator, a physiotherapist with 5 years of experience in level one musculoskeletal real-time ultrasonography, screened potential participants for inclusion and exclusion criteria, recorded participant data and demographics, and conducted all RTUI and TrA measurements.

8.5.4 Experimental equipment, outcome measures and procedure

Standard protocols were used to measure height with a stadiometer and weight with weighing scales (Norton & Olds, 2006). RTUI (GE Healthcare Venue 40 MSK; General Electric Company; Wauwatosa, WI, USA) was conducted in movie mode, using a 3.1 MhZ curved-array abdominal probe 4C-SC model 5337596 (65 × 15 mm footprint), to capture real-time video images of the participants' dominant side RTrA and CTrA over two separate measurement sessions ('Measurement 1' and 'Measurement 2'). An FPD was attached to the RTUI probe (see Plate 8.1), and real-time on-screen display of probe force (N), inclination (°) and roll (°) was recorded at 60 Hz via a LabVIEW virtual instrument link on a laptop computer (Gilbertson & Anthony, 2013), and stored for later analysis (see Plate 8.2).

During imaging, the RTUI machine, FPD and examiner were positioned to the right of participants, who lay on a surgical procedure bed in the supine 'crook' position, with no pillow head support (Mannion et al., 2008). Pelvic and lumbar

152

position were standardised by palpation and auditory cueing (O'Sullivan, 2005).

Goniometry was used to standardise hip and knee joint angles at 30° and 90°,

respectively. A breathing cycle protocol was used for image capture of RTrA and

CTrA (Whittaker, Warner, & Stokes, 2010; see Table 8.1).

      A familiarisation session was conducted prior to RTUI. For RTrA,

participants were instructed to breathe normally and to then hold the end-expiratory

phase for up to 3 s. For CTrA, participants were taught an abdominal draw-in

manoeuver (ADIM; Whittaker et al., 2010) to selectively activate the TrA

(Richardson, Hodges, & Hides, 2004). Initial ADIMs were conducted at maximum

effort for familiarisation but, during CTrA RTUI, intensity was standardised to

ADIMs conducted at 50% effort (Henry & Westervelt, 2005; Whittaker et al., 2010).

      The RTUI probe was clipped into the FPD shell, and a water-based coupling

gel was applied to the probe footprint. A sheath (14×121.9 cm; 3D end with

PullUp™; Protek Medical Products Inc., Coralville, IA) was placed over the probe

and shell, then secured in place with an elastic restraining band (see Plate 8.1). In

preparation for RTUI, the FPD software and the RTUI machine were activated, then

the probe cable underwent 'strain relieving' and force bias 'zeroing' was conducted

(Gilbertson & Anthony, 2013). No skin shaving was performed. A layer of coupling

gel was applied to the sheath surface, and the probe was placed lateral to the

umbilicus, midway between the iliac crest and the lower ribs (Costa, Maher, Latimer,

Hodges, & Shirley, 2009).

      Preliminary RTUI established the probe position required to achieve optimal

RTrA and CTrA views, and this position was marked with a Hypafix® (BSN Medical

Luxembourg Finance Holding S.à r.l., Luxembourg) adhesive template applied to the

participants' skin (see Plate 8.1). For Measurement 1, RTUI was synchronised with

the FPD software. For RTrA, minima and maxima probe force, inclination and roll were recorded from the on-screen display (see Plate 8.2) over 2–3 full breathing cycles (see Table 8.1). CTrA was conducted using the same methods for at least two

Plate 8.1. The 'force probe device' attached to the real-time ultrasound probe and standardised position using adhesive template.
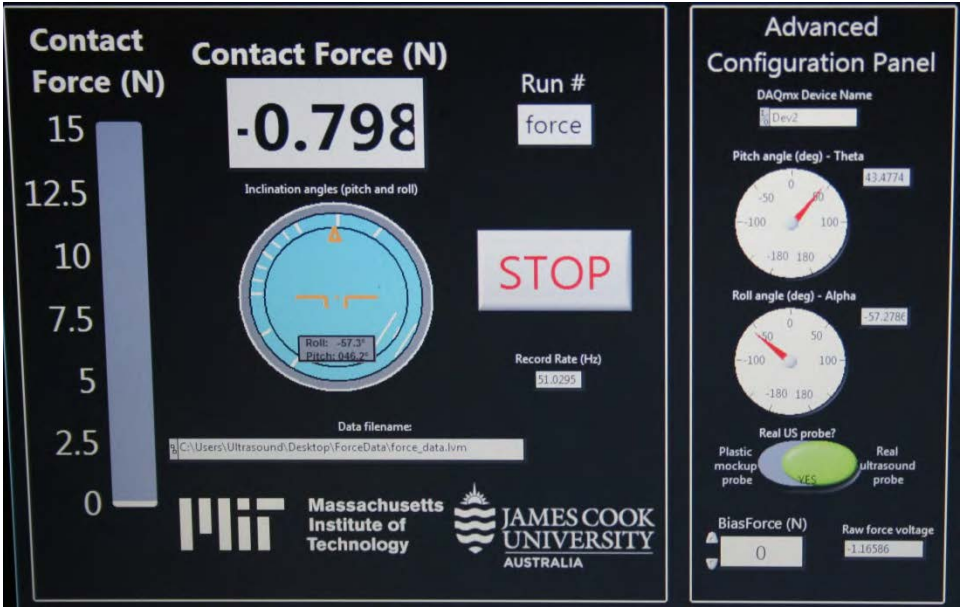
Plate 8.2. 'Force probe device' on-screen display.

ADIM cycles (see Table 8.1). Participants then rested for 15 min. Thereafter, participants and probe position were re-established and the adhesive template was reapplied. Measurement 2 was conducted, replicating the methods of Measurement 1. Ranges of probe force, inclination and roll were standardised across the two separate measurements and between RTrA and CTrA. To avoid ADIM 'training bias', RTUI was visible only to the examiner, with participants blinded. RTUI and FPD data were automatically stored to a memory card for later processing and analysis.

8.5.5 Still image extraction

Data collection yielded one RTrA and one CTrA video for Measurement 1 and for Measurement 2 per participant ($N = 4$), from which cropped video sections of RTrA during the end-expiratory phase and of CTrA during ADIM for each measurement (stage 3 of the breathing cycle protocol; see Table 8.1) were extracted and stored ($N = 68$). Cropped videos for each participant were time matched to the FPD data. Microsoft Excel™ was used to sort data to identify FPD data time points where probe force, inclination and roll matched. The matched data time points were identified on each video, and four still RTUI images per participant with matched probe force, inclination and roll were acquired ($N = 68$).

To avoid measurement bias, an independent research assistant de-identified each image using a computer-generated random identity, thus blinding the examiner prior to measurement. ImageJ measurement software (Schneider, Rasband, & Eliceiri, 2012) was used to calibrate image size. TrA thickness (mm) was then measured as the perpendicular distance between the inside margin of its upper fascial borders, taken from a point 25 mm from the inside edge of the medial fascial join (see Plate 8.3). Blinding was then removed and images re-identified. For each pair of copied images, measurements were averaged and reported as 'mean RTrA' or 'mean

Table 8.1

A Single Breathing Cycle Protocol During Real-Time Ultrasound Imaging

Stage	Video phase	Imaging of resting transversus abdominis		Imaging of contracted transversus abdominis during abdominal draw-in manoeuvre	
		Verbal Instruction	Process	Instruction	Process
1	Inspiratory phase	'Imaging will start as you begin your next breath in.'	Start imaging and force probe data collection simultaneously.	'Imaging will start as you begin your next breath in.'	Start imaging and force probe data collection simultaneously.
2	Expiratory phase	'Breathe out normally.'	Exhale.	'Visualise an imaginary line between your pelvic bones. Breathe out normally, gently and slowly tighten your abdomen and draw the imaginary line towards your spine. Keep your lower back still.'	Exhale and conduct transversus abdominis submaximal contraction.
3	End of expiratory phase	'Hold the breath out now remaining relaxed, 1, 2, 3.'	Maintenance of exhalation, transversus abdominis relaxed.	'Hold the breath out now, keeping your abdomen drawn towards your spine and lower back still, 1, 2, 3.'	Maintain transversus abdominis submaximal contraction.
4	Inspiratory phase	'Breathe in normally.'	Inhale.	'Breathe in normally and slowly relax your abdomen.'	End of abdominal draw-in manoeuvre.

CTrA'. TrA-C was represented as the 'mean RTrA' to 'mean CTrA' change in TrA thickness using the formula (CTrA thickness − RTrA thickness/RTrA thickness; Koppenhaver, Hebert, Fritz, et al., 2009).

8.5.6 Data analysis

Statistical analysis was conducted using SPSS software, version 22 (SPSS Inc., Chicago, IL). Descriptive statistics were generated for gender, age and BMI as a measure of participant's relative size, calculated from height and weight measurements (kg/m2; Dahl et al., 2013). The FPD has a 6-axis force/torque measurement expressed in X, Y and Z axes (Gilbertson & Anthony, 2013); thus, negative and positive values of inclination and roll were represented using the right-hand rule for reporting of kinematic data (Wu & Cavanagh, 1995). Ranges of probe force, inclination and roll, and numerical variables for RTrA, CTrA and TrA-C, were analysed for normality of distribution and reported using minimum, maximum, mean and *SD*, or *Mdn* with *IQR*, as appropriate.

$ICC_{3,2}$ with 95% CI were calculated for the two mean repeated imaging measurements where participants were repositioned on the procedure bed, and both participant and FPD were repositioned between RTUI measurements. Measurement precision was reported as the standards error of measurement (*SEM)*, calculated using the formula $SDx\sqrt{1-ICC}$ (Portney & Watkins, 2009).

8.6 Results

Descriptive statistics for the characteristics of the study population (N = 17) are summarised in Table 8.2. All still images were of good quality and measurable. Each still image was copied, yielding 68 duplicate paired images. This pilot study achieved a post hoc statistical power of .98. Probe force, inclination and roll during RTUI ranged from 0.88 to 5.26 N (2.99 ± 0.99), from −19.85° to 59.21° (3.92 ±
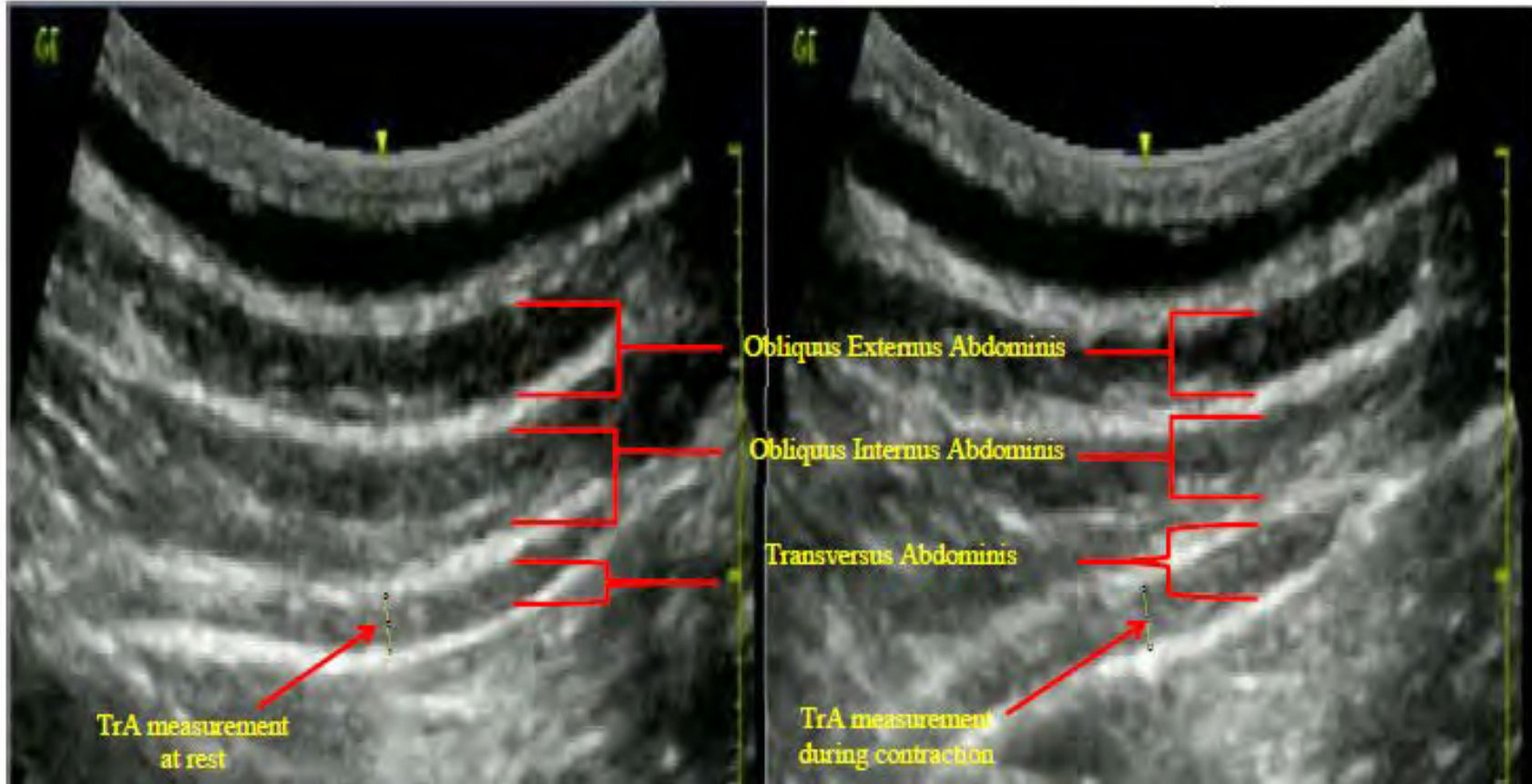
Plate 8.3. Images showing measurement of transversus abdominis during rest and abdominal draw-in manoeuvre

16.57), and from −41.72° to 62.84° (22.46 ± 39.11), respectively. RTrA (*ICC* = .97, CI [0.93, 0.99]); CTrA (*ICC* = .99, CI [0.97, 0.99]); and TrA-C (*ICC* = .93, CI [0.81, 0.97]) measurements were highly reliable (Shrout & Fleiss, 1979; see Table 8.3).

8.7 Discussion

This study investigated the utility of an FPD to determine a standardised method for TrA RTUI, and reported a single examiner's measurement reliability of RTrA, CTrA and TrA-C in a CLBP population using the standardised method. Only two prior TrA RTUI studies have reported examiner reliability specifically in CLBP (Costa, Maher, Latimer, Hodges, & Shirley, 2009; Mannion et al., 2008). *ICC*s here ranged from .32 to .72.

Table 8.2

Descriptive Characteristics

	All (*N* = 17)	Male (*N* = 13)	Female (*N* = 4)
Age (yr)	49.8 ± 13.1	48.4 ± 14.5	54.25 ± 7.14
Height (m)	174.64 ± 6.39	176.46 ± 5.44	168.75 ± 6.24
Mass (kg)	84.41 ± 16.51	85.64 ± 13.54	80.38 ± 26.31
BMI (kg/m^2)	27.67 ± 5.31	27.46 ± 3.79	28.34 ± 9.6

Note. Values are mean ± *SD*.

Table 8.3

Intra-Examiner Reliability

	ICC	95% CI	*SEM*
Resting transversus abdominis	.98	[0.93, 0.99]	0.1
Contracted transversus abdominis	.99	[0.97, 0.99]	0.1
Transversus abdominis activation (%)	.93	[0.82, 0.97]	0.09

Note. CI = confidence interval; *ICC* = intra-class correlation coefficient; *SEM* = standard error of measurement.

The results of the current study indicated that dominant-sided TrA thickness measurements were highly reliable for RTrA (*ICC* = .97, CI [0.93, 0.99]) and CTrA (*ICC* = .99, CI [0.97, 0.99]). TrA-C was also highly reliable (*ICC* = .93, CI [0.81, 0.97]). Such results therefore appeared superior to those previously obtained.

The results endorse the use of real-time feedback to ensure consistency in terms of probe force and orientation throughout TrA RTUI. Such consistency permitted the comparison of images obtained at two measurement sessions, each for RTrA and CTrA, with probe force, inclination and roll matched across four images. Thus, probe parameters were controlled for during TrA measurement. Such control has not been realised in 'free-hand' RTUI. Even in one study where RTUI probes were secured to the patient using foam and a belt, TrA-C *ICC* values of only .62 were obtained (Mannion et al., 2008).

However, other factors could have influenced the present study results. For example, in contrast to the current study, Costa, Maher, Latimer, Hodges and Shirley (2009) utilised an involuntary CTrA via isometric knee movement. Further, the training status of the examiner is important in any study of examiner agreement. In this study, the examiner was a physiotherapist with 5 years' experience in level one musculoskeletal RTUI. While examiner training level was mentioned in the TrA-C

160

CLBP study by Costa, Maher, Latimer, Hodges and Shirley (2009), this was not mentioned in that of Mannion et al. (2008). The extent to which differences in the examiner experience might have affected the results reported herein cannot be assumed. Although this study did not compare examiners of various experience levels, the results reported suggest that using the FPD has potential to improve intra-examiner reliability regardless of experience level.

This study is the first to produce standardising ranges of force, inclination and roll using an FPD during TrA RTUI in CLBP patients. The ranges observed in the current study reflect the residual variation still apparent in probe force and orientation despite the advantage of visual feedback associated with FPD RTUI. Optimal *ICC*s for RTrA, CTrA and TrA-C were obtained from standardised ranges of probe force, inclination and roll in the current study. It is likely that greater ranges would be observed during 'free-hand' RTUI, and that these would account for the lower *ICC*s observed in prior studies.

Participants in this CLBP study had an 'overweight' BMI, with a mean age over 49 years. This population presented TrA RTUI challenges associated with higher adiposity and anatomical landmark identification, with potential compromise to the results found when using the FPD. However, despite these challenges, the reliability reported in this study was high, supporting the application of these methods in the future for longitudinal research.

There were limitations to this study, the FPD equipment is a recent development and although the bioengineering technology of the device is established to standardise force, inclination and roll of the probe during RTUI, it is currently not integrated into any commercially available RTUI machines, limiting its availability for general clinical application. A further limitation relates to the fact that only intra-

examiner reliability was assessed. Additionally, the reliability of the examiner using 'free hand' RTUI compared to FPD RTUI was not compared. Therefore, future studies are required to assess intra- and inter-examiner reliability, and to compare 'free hand' with FPD RTUI reliability. Further, only dominant-side TrA examinations were performed. Finally, male subjects outnumbered female subjects in this convenience pilot study of CLBP patients. Notwithstanding, 68 pairs of duplicate images were obtained for analysis. Blinding and randomisation reduced measurement bias.

8.8 Conclusion

This study established the utility of an FPD to achieve highly reliable intra-examiner RTUI TrA measurement in CLBP patients. Results obtained were superior to prior reports using free-hand RTUI. Unique standardising ranges, with mean values, for probe force, inclination and roll were obtained to aid future studies and research. FPD RTUI is likely to optimise TrA-C assessment in CLBP populations and indeed other populations.

Key points:

- Free-hand RTUI TrA measurement reliability in CLBP is suboptimal with poor to moderate *ICC*s.

- New technology known as an FPD was available to standardise RTUI probe force, inclination and roll.

- The FPD technology has utility for RTUI TrA in CLBP patients, but requires evaluation for TrA measurement reliability.

- A standardised method of image acquisition and interpretation was demonstrated for RTUI TrA measurement using the FPD in CLBP.

- Highly reliable intra-examiner RTUI TrA measurement was achieved for the standardised method using the FPD in CLBP.

- Reported reliability was superior to prior reports using free-hand RTUI.

- These results support the use of this RTUI TrA methodology for the succeeding study (Chapter 9), to address the fifth research question of this thesis (see Chapter 1).

8.9 How This Chapter Informed Subsequent Stages of the Research

It was important to this research that the methods used to conduct the final study (see Chapter 9) were rigorous, and the physiotherapist's RTUI TrA measurement demonstrated high reliability. This chapter outlined the confounding factors that result in the poor to moderate levels of RTUI TrA measurement reliability reported previously in CLBP studies. This chapter also explained the proposed methods of TrA RTUI using the FPD and data extraction process. Further, the high intra-examiner reliability of repeated measurements reported in this study supported the utility of the FPD to optimise the proposed methodology for the subsequent study of this thesis (see Chapter 9).

Chapter 9: Activation of the Transversus Abdominis in Chronic Low Back Pain: The Immediate Effect of Pain Abolition

This chapter addresses the sixth research question (see Chapter 1): 'What is the immediate effect of pain relief on TrA activation in a CLBP population?'

9.1 Introduction

A plethora of research has contributed to understanding the role of the TrA in the clinical presentation of LBP. Significantly reduced transversus abdominis activation (TrA-C) has been reported in CLBP, compared with asymptomatic individuals (Critchley & Coutts, 2002; Kim et al., 2013; Pulkovski et al., 2012), and it has been suggested that this may contribute to ongoing disability (Hodges & Moseley, 2003). Rehabilitation of the transversus abdominis (TrA) to improve spinal stability and thus limit symptoms in CLBP has been advocated (Richardson & Jull, 1995). This research has influenced contemporary physiotherapy practice, with RTUI used as a tool to relay on-screen visual feedback of TrA activation to educate and retrain TrA function in symptomatic individuals (Anderson Worth, Henry, & Bunn, 2007; Chipchase, Thoirs, & Jedrzejczak, 2009).

Greater TrA-C measured by thickness changes with real-time ultrasound imaging (RTUI) indicates improved function (Koppenhaver, Hebert, Parent, & Fritz, 2009; Lariviere et al., 2013). TrA-C has been shown to reduce in response to experimentally induced LBP (Hodges et al., 2003; Kiesel et al., 2008). Also, reduced postural activation of TrA-C due to anticipation of experimental back pain has been reported (Moseley, Nicholas, & Hodges, 2004). In contrast, there is evidence to suggest that neuromuscular changes seen in CLBP are independent of pain, and may

be dependent on more complex interactions between psychological and neuromusculoskeletal factors (Dubois et al., 2014). Thus, evidence suggests that in CLBP TrA-C dysfunction may not be the result of pain inhibition alone (Dubois et al., 2014; Tsao et al., 2011).

The objective of this study was to assess the effect of pain abolition on TrA-C in a CLBP population, using a standardised RTUI method, conducted before and immediately after pain abolition by diagnostic spinal anaesthetic injection (DSAI) . It was hypothesised that due to complex neuro-musculo-skeletal factors present in CLBP, that immediate pain relief from a DSAI would result in no statistical or clinically significant difference in pre–post TrA-C. This study would provide evidence about inherent TrA-C response to sudden abolition of CLBP.

9.2 Method

9.2.1 Study design

This pre–post intervention study was part of a wider study registered with the Australian New Zealand Clinical Trials Registry on 6 March 2013 (ACTRN: 12613000267752).

9.2.2 Study location

The study was conducted in a neurosurgeon's specialist back pain spinal injection clinic at two secondary-care day procedure facilities in North Queensland, Australia. Participants were referred to the clinics by either a primary healthcare physician or a healthcare specialist.

9.2.3 Ethics

Study approval was obtained from the Human Research Ethics Committees of James Cook University (H4387), Townsville Hospital (HREC10QTHS53; see Appendix 4B) and Mater Health Services (MHS20150512-07; see Appendix 4C). An

explanation and information sheet detailing aims and methods of the study was provided to participants, and written consent was gained prior to study commencement.

9.2.4 Participants and examiner

Potential participants were consecutively approached as they attended the clinics between August and October 2015. The primary researcher, an experienced physiotherapist with training in RTUI, screened potential participants for inclusion and exclusion criteria, collected demographic and anthropometric data, and conducted all RTUI and TrA measurements. A consultant neurosurgeon, also trained in orthopaedic spine surgery, with over 20 years of experience in CLBP management and performing spinal injection procedures, examined participants and, if clinically indicated, conducted DSAI in accordance with medical guidelines and routine hospital procedures.

9.2.5 Inclusion and exclusion criteria

The inclusion and exclusion criteria applied in this study have been detailed previously in Chapter 8.5.2 of this thesis.

9.2.6 Study equipment and outcome measures

Participant age (yr), gender, weight (kg), height (m), hand dominance, and symptom intensity, duration and distribution were recorded by the physiotherapist. Self-reported pain intensity was measured using an 11-point millennium VAS (Huskisson, 1974). The VAS was chosen as it has greater reliability and sensitivity for reporting pain, compared with other measures (Hawker et al., 2011; Huskisson, 1974). Pre- and post-VAS measurements were used to establish percentage pain reduction immediately following DSAI.

A real-time ultrasound scanner, details of which have been reported previously (see Chapter 8.5.4), was used to capture real-time video images of the transversus abdominis at rest (RTrA) and submaximal transversus abdominis contracted (CTrA) bilaterally. A force probe device (FPD) (Gilbertson & Anthony, 2013), previously detailed in Chapter 8, was attached to the real-time ultrasound probe during video capture. For methodological details, FPD display and data recording, please refer to Chapter 8.5.4.

9.2.7 Sample size estimation

A sample of 50 participants was required to reach 80% power, assuming a type 1 error probability of 5%, an estimated mean pre- to post-DSAI thickness (mm) change ratio of 1.73:l, the ability to detect a minimum of 10% change in TrA activation, and factoring for dropout due to poor image quality or participants with an immediate pre–post DSAI reduction in the VAS score of less than 75%.

9.2.8 Reliability of imaging and still image measurement

High reliability of the examiner to measure TrA-C in CLBP was established prior to conducting the study (see Chapter 8.6).

9.2.9 Preliminary study procedures

Participants were prepared for the DSAI procedure according to the standard hospital protocols. Participants remained on the same surgical procedure bed for pre-procedure RTUI, DSAI and post-procedure RTUI (see Plate 9.1). The RTUI machine with attached FPD, and examiner and patient positioning for each RTUI capture, has been detailed previously (see Chapter 8.5.4). Participants unable to replicate lower limb angles because of reduced joint flexibility were positioned with joint angles as close as possible to those preferred, and these measured angles were replicated exactly for each RTUI capture.

The following aspects of this study methodology were applied as described previously (see Chapter 8.5.4):

- pre-RTUI familiarisation session

- breathing cycle protocol for RTrA and CTrA capture

- process for conducting a submaximal CTrA

- attachment of the FPD to the RTUI probe, and preliminary set-up

- activation of the RTUI machine and FPD software

- participant skin preparation

- preliminary RTUI probe-to-skin positioning.

Plate 9.1. Surgical bed for diagnostic spinal anaesthetic injection procedures.

9.2.10 Pre-and post-fluoroscopy-guided diagnostic spinal anaesthetic injection image acquisition

The RTUI machine and FPD software were synchronised, and a video with a duration of 20–40 s of the right RTrA was captured while the participant completed two to three full breathing cycles according to processes detailed previously in

Chapter 8 (see Table 8.1). Simultaneously, minima and maxima force, inclination and roll FPD parameters were monitored by the examiner via the laptop display, which were recorded and stored via the FPD software. Pre-DSAI imaging of the right CTrA followed, comprising two or three consecutive abdominal draw in manoeuvres (ADIMs) captured over a 20–40 s duration. The physiotherapist maintained the same minima and maxima force, inclination and roll FPD parameters for each RTUI capture. This same process was repeated for the left RTrA, followed by CTrA RTUI. This completed all pre-DSAI imaging procedures. To avoid training bias, displayed images were visible to the physiotherapist but the participant was blinded during RTUI capture.

Post-DSAI, the same imaging methods were applied, with the replicated minima and maxima FPD parameters maintained across all videos. The RTUI videos and associated FPD data were stored to a memory card for later processing on a personal computer.

9.2.11 Fluoroscopy-guided diagnostic spinal anaesthetic injection procedures

The neurosurgeon conducted the DSAI immediately following pre-procedure imaging, according to methods previously described (see Chapter 7.2.8).

9.2.12 Still real-time ultrasound image extraction and measurement

Data collection yielded four pre- and four post-DSAI rest and contraction paired videos per participant ($N = 8$). Poor quality immeasurable videos were identified and excluded. Still image extraction was completed according to methods previously described (see Chapter 8.5.5). Two still images were captured from each video, providing all videos were of good quality and measurable. Thus, eight pre-DSAI and eight post-DSAI still images were available for each participant. A total of

752 still images (376 pre-DSAI and 376 post-DSAI still images for all participants) were expected following image extraction.

Still image blinding and image measurement with ImageJ software (Rasband, 1997) were conducted according to methods described previously (see Chapter 8.5.5). A mean TrA thickness (mm) was calculated from each set of two still images captured from each video. Figure 9.2 illustrates the flow of still image processing to achieve TrA-C measurement. 'Image 1' and 'Image 2' represent the images captured for each of the resting and contracting states for the right and left sides, each 'Image' representing the mean measurement of each set of two still images captured from each video.

The formula described previously (see Chapter 8.5.5) was used to calculate 'Image 1' and 'Image 2' TrA-C from 'Image 1' and 'Image 2' RTrA and CTrA measurements bilaterally (see Figure 9.2). The final right and left TrA-C represent the mean of 'Image 1' and 'Image 2' TrA-C (see Figure 9.2).

9.2.13 Data analysis

Statistical analysis was conducted in SPSS software, version 22 (SPSS Inc., Chicago, IL). The percentage change in the VAS scores from pre- to post-DSAI was calculated for each participant. Pain abolition was defined as a minimum of 75% reduction from the baseline pre-DSAI VAS score. BMI was calculated as a measure of participant's relative size (kg/m^2) taken from height and weight measurements (Dahl et al., 2013).

Descriptive statistics for gender, symptom duration, age, BMI and baseline VAS score were generated, analysed for normality of distribution and reported using mean and standard deviation (*SD)* ($\overline{x} \pm SD$), or median (*Mdn)* with inter-quartile range (*IQR)*, as appropriate. RTUI probe parameters of force, inclination and roll

were generated from FPD data output and ranges reported. Negative or positive

values of inclination and roll were reported using the right-hand rule for reporting of

kinematic data (Wu & Cavanagh, 1995).

```
┌─────────┐                              ┌─────────┐
│  Right  │                              │  Left   │
└─────────┘                              └─────────┘
```

| Image 1: Rest measure (mm) | Image 2: Rest measure (mm) | Image 1: Rest measure (mm) | Image 2: Rest measure (mm) |

| Image 1: Contraction measure (mm) | Image 2: Contraction measure (mm) | Image 1: Contraction measure (mm) | Image 2: Contraction measure (mm) |

| Image 1 TrA-C: Rest to Contraction % change | Image 2 TrA-C: Rest to Contraction % change | Image 1 TrA-C: Rest to Contraction % change | Image 2 TrA-C: Rest to Contraction % change |

| Right mean TrA-C ($N = 43$) | Left mean TrA-C ($N = 41$) |

Figure 9.2. Flow of still image processing to measure transversus abdominis

activation.

To assess for potential learning or fatigue effect from repeated ADIMs, a

preliminary analysis was conducted. 'Image 1' and 'Image 2' CTrA measurements

were analysed for normality of distribution and a repeated-measures statistical

analysis was conducted to assess for significant differences ($p < .05$) between CTrA

'Image 1' and 'Image 2' measurements bilaterally, pre-DSAI and post-DSAI. Any differences were located with pairwise Bonferroni corrections.

To assess the effect of pain abolition on TrA-C, numerical variables of TrA-C were analysed for normality of distribution, and appropriate paired bivariate statistical tests conducted to assess for statistically significant differences ($p < .05$) in pre- to post-DSAI TrA-C of the study sample, with comparison between participants who reached the stated pain abolition cut off level and those who did not. Clinically significant differences in pre- to post-DSAI TrA-C were also assessed, with a 20% or greater TrA-C pre–post DSAI change considered appropriate on the basis of a previous study where experimentally induced LBP resulted in 20% lower TrA-C, compared with a control group (Kiesel et al., 2006).

Bivariate independent statistical tests were conducted to assess for significant pre- to post-DSAI TrA-C differences ($p < .05$) in gender, hand dominance, pre-DSAI pain intensity, duration of symptoms and side of symptom categories.

9.3 Results

9.3.1 Study sample

All patients who attended the weekly spinal injection clinics ($N = 58$) were invited to the study. A consecutive convenience sample of 30 male and 17 female subjects were recruited ($N = 47$, 81% of invited participants).

9.3.2 Pain intensity response to fluoroscopy-guided diagnostic spinal anaesthetic injection

CLBP was totally abolished immediately following DSAI in 92% of participants. The remainder reported a 99% reduction in the VAS score post-DSAI.

9.3.3 'Force probe device' parameter ranges

During RTUI, the FPD-recorded probe parameters ranged from 0.88 to 11.44 N ($mean = 3.26$ N) for force, from $-48.16°$ to $110.98°$ ($mean = 3.15°$) for inclination and from $-70.18°$ to $76.99°$ ($mean = 3.69°$) for roll.

9.3.4 Still images

Two still images extracted from each RTUI video resulted in an expected total of 376 pairs of still images for the study measurements (16 per participant). However, 52 pairs (14%) were of poor quality and immeasurable and were subsequently excluded, resulting in 324 RTrA to CTrA pairs of still images to be measured.

9.3.5 Learning or fatigue effect on contracted transversus abdominis thickness with repeated abdominal draw-in manoeuvres

'Image 1' and 'Image 2' CTrA measurements were normally distributed. A repeated-measures ANOVA was conducted to rule out potential learning effects or fatigue on CTrA over repeated ADIMs during RTUI. Results showed no significant differences between 'Image 1' and 'Image 2' CTrA measurements bilaterally either pre-DSAI (F = .01, p = .93) or post-DSAI (F = .57, p = .46). This eliminated either a learning or fatigue effect from repeated ADIM as a confounder in this study.

9.3.6 Demographic and baseline transversus abdominis activation characteristics

Descriptive statistics for the characteristics of the study population are summarised in Table 9.1. Participant age and BMI were normally distributed, while symptom duration and baseline VAS score were not normally distributed. Male subjects ($N = 30$), who represented 64% of the population, were significantly younger and had a significantly lower baseline pain intensity than did female

participants ($N = 17$). There was no significant difference between male and female subjects in body mass index (BMI) and symptom duration (see Table 9.1). Additionally, pre-DSAI TrA-C did not differ significantly between male and female subjects (Z = -0.047, p = .83, CI [−0.4, 0.25]).

9.3.7 Change in transversus abdominis activation immediately following diagnostic spinal anaesthetic injection

A Wilcoxon signed-rank test showed no significant difference between pre- and post-DSAI TrA-C (Z = 0.03, p = .66, CI [−0.13, 0.18]). No significant differences between pre- and post-DSAI TrA-C in male (Z = -0.02, p = .78, CI [−0.17, 0.13]) and female (Z = 0.05, p = .70, CI [−0.20, 0.30]) participants existed. Neither was there a clinically significant difference (20%) between pre- and post-DSAI TrA-C ($Mdn = 3\%$, SD 47%).

A Mann–Whitney test conducted for between-gender comparison of pre-post-DSAI TrA-C indicated no significant differences between males and females (Z = -0.1, p = .64, CI [-0.3, 0.2]). Similarly, no significant differences in pre- to post-DSAI TrA-C were identified when dominant and non-dominant hand were compared, or when symptomatic and asymptomatic side of the body were compared (see Table 9.2).

Kruskal–Wallis tests showed no significant differences in pre- to post-DSAI TrA-C when pre-DSAI pain intensity or symptom duration categories were compared (see Table 9.2).

Table 9.1

Descriptive Data

	Total (N = 47)	Male (N = 30)	Female (N = 17)	p value [CI]
Age ($\bar{x} \pm SD$)	50.53 ± 16.02	45.43 ± 15.51	59.53 ± 12.94	.001* [−23.05, −5.14]
BMI (kg/m^2) ($\bar{x} \pm SD$)	28.35 ± 5.09	28.22 ± 3.94	28.66 ± 7.01	.82 [−4.53, 3.63]
Symptom duration (months) *Mdn* (*IQR*)	24 (111.25)	24 (82)	48 (106.5)	.32 [−42, 11]
Baseline VAS *Mdn* (*IQR*)	45.5 (39.25)	31(30.5)	55 (22)	.002* [−36, −9]

Note. VAS = visual analogue scale.

*Indicates significance at $p < 0.05$.

9.4 Discussion

This study aimed to identify the effect of immediate pain relief on TrA function. The findings of this study indicate that a 99% or greater reduction in baseline pain intensity experienced by participants had no statistically or clinically significant effect on their ability to activate the TrA. Additionally, gender, hand dominance, side of CLBP, pre-procedure pain intensity and duration of symptoms had no significant effect on their ability to activate the TrA following a pain-relieving DSAI. This indicates that pain inhibition has minimal effect on TrA-C in a CLBP population, thus supporting the hypothesis that the complex interaction of prolonged symptoms, neuromuscular deficit and other aetiological factors would prevent immediate improvement in TrA-C following pain reduction or abolition.

Pain reduction and restoration of functional loss are fundamental aims of physiotherapy CLBP management. Hence, an understanding of the effect of pain

Table 9.2

Group Comparisons of Transversus Abdominis Activation Pre- to Post-Diagnostic Spinal Anaesthetic Injection

		Transversus abdominis activation pre- to post-diagnostic anaesthetic spinal injection (%)	*p* value [CI]
Hand dominance[a]	Dominant	0.08 (0.02)	.26 [−0.2, 0.45]
	Non-dominant	−0.18 (−0.04)	
Symptom location^	Symptomatic side	−0.03 (−0.06)	.81 [−0.35, 0.35]
	Asymptomatic side	0.1 (0.0)	
Baseline VAS^	Mild (>5–44 mm)	0 (−0.08)	
	Moderate (45–74 mm)	−0.05 (0.08)	.49
	Severe (75–100 mm)	−0.4 (0.9)	
Symptom duration^	3–12 months	0.1 (0)	
	13–60 months	0 (−0.05)	.95
	61–120 months	−0.15 (0.15)	
	>120 months	0.2 (−0.3)	

Note. CI = confidence interval; VAS = visual analogue scale.
[a]Reported as mean ± *SD*; [b]Reported as median (*IQR*).

177

abolition on TrA-C is important to establish efficacy and appropriate timing of physiotherapy TrA facilitation interventions. However, in isolation, pain reduction only addresses one aspect of the CLBP presentation.

Although DSAI is a pain-relieving procedure in CLBP medical management (Wu, Zhao, Dong, Song, & Li, 2016), in isolation the pain-relieving effect does not provide a holistic approach to CLBP treatment. Specifically, this single intervention does not address potential neuromusculoskeletal contributions to patients' clinical presentation and reduced quality of life. This was demonstrated by the outcome of the current study, which suggests that a pain-relieving intervention without subsequent TrA-C retraining is unsatisfactory.

This study has provided evidence that immediate pain relief does not spontaneously improve TrA-C in CLBP. This supports previous research in a heterogeneous LBP study sample, which suggested that pain inhibition is not solely responsible for deficient activation of another spinal stabilising muscle, the lumbar multifidus (Hides, Richardson, & Jull, 1996). Indeed, retraining TrA-C concomitant with lumbar multifidus activation, to achieve optimal spinal stability, has previously been considered a prerequisite physiotherapy management strategy in patients identified with TrA deficit in CLBP (Hides, Stanton, Dilani Mendis, & Sexton, 2011). Further, others have suggested that pain relief achieved from lumbar facet injections should be succeeded by physiotherapy treatments, which may improve long-term outcomes in CLBP (Chambers, 2013). However, to the author's knowledge, no prior study has provided evidence to directly support the need for physiotherapy TrA-C retraining immediately following CLBP abolition. Therefore, the findings of the current study have substantial clinical significance as they indicate

that physiotherapy-specific TrA-C retraining for CLBP patients may be efficacious following pain-relieving procedures.

It is acknowledged that this study did not identify how long after DSAI it is optimal to commence TrA-C retraining. Thus, prospective studies are warranted to determine if there is an optimal time to begin TrA-C-specific therapeutic retraining, and what is the most desirable TrA-C facilitation activity in CLBP. Further, studies at progressive intervals post-DSAI, to determine if a TrA-C 'refractory period' exists and its duration, are indicated.

A limitation of this study relates to the role of fear avoidance after pain abolition. This study did not explore the effect of post-DSAI fear avoidance on TrA-C. Fear avoidance post–pain relief may be considered a potential confounder, and no assumptions can therefore be made as to whether fear avoidance immediately post–pain abolition affected TrA-C in this study sample. Additionally, male subjects outnumbered female subjects in this study population of CLBP patients. Although, participants were consecutively recruited to the current study, the gender distribution differed from the study population reported in Chapter 7. Unlike the study reported in Chapter 7, the current study was conducted across two hospital sites. This might have affected the male-to-female recruitment ratio. Specifically, the gender ratios of patients who attended the clinics and required DSAI, and of those who volunteered, might have differed.

Despite some limitations, this study has demonstrated superior methodology, compared with earlier free-hand TrA RTUI studies. The FPD technology method used in this research was rigorous and highly reliable. FPD equipment is unique, as it standardises force, inclination and roll of the probe during RTUI, resulting in methodological standardisation, which is not possible during the uncontrolled free-

hand TrA RTUI applied in research to date. As such measurement bias in this study was minimal, it is recommended that the methods described herein are appropriate for longitudinal and repeated RTUI TrA measurements, and should be adopted in future studies.

9.5 Conclusion

Using rigorous and standardised TrA RTUI methods, this study established that, in a consecutive convenience sample of CLBP patients, immediate pain relief from DSAI did not significantly affect TrA-C. Results obtained suggest that pain relief as an isolated intervention is insufficient to address the complex interaction of factors that contribute to TrA-C dysfunction in CLBP. Results provide foundation knowledge to support efficacy and appropriate timing of physiotherapy TrA-C facilitation interventions for CLBP patients.

Key points:

- Significantly reduced TrA-C has been reported in CLBP patients compared with asymptomatic individuals.

- Previously, the immediate effect on TrA-C following CLBP reduction or abolition has not been reported.

- Results showed that a 99% or greater reduction in baseline pain intensity experienced by participants had no immediate effect on their ability to activate the TrA, however it is unclear if sustained reduction in pain would result in increased activation of TrA over time.

- In this study, rigorous and standardised TrA RTUI methods using an FPD to standardise probe force, inclination and roll were used to report differences in pre- to post-DSAI TrA-C.

- Unlike previous studies using free-hand TrA RTUI, the results of this study are highly reliable, as they were subject to minimal measurement bias, and should be considered whenever longitudinal data are required.

9.6 What This Chapter Adds to Current Clinical Knowledge

It was important to report the effect of immediate pain relief on TrA-C in CLBP, as this had not been conducted previously. The findings of this chapter suggest that the absence of pain does not automatically improve TrA function. However it is unclear if sustained reduction in pain would result in increased activation of TrA over time. This study highlights the importance of multidisciplinary management strategies for CLBP patients.

Chapter 10: Outcomes, Significance and Clinical Implications, Future Research Recommendations, and Conclusions

10.1 Research Outcomes with Respect to the Research Objectives

10.1.1 Objective 1

To identify via systematic review valid and reliable physiotherapy LBP classification systems and clinical examination tests with high diagnostic accuracy for CLBP populations.

Two systematic reviews indicated a paucity of research about the concurrent validity, diagnostic accuracy and reliability of physiotherapy LBP classification systems and clinical examination tests when applied to CLBP populations. The evidence was insufficient to support the clinical application of any existing LBP classification system as a valid and reliable standalone examination process for CLBP. However, existing research supported the inclusion of some groups or clusters of examination techniques to classify CLBP.

Five studies reported validity or diagnostic accuracy of three groups of clinical examination tests or clusters in CLBP. Overall, these studies demonstrated a moderate risk of bias. Three additional studies reported inter-rater reliability of two lumbar spine classification systems in CLBP. Reliability was variable ($k = .32–.96$), with an overall high risk of study bias.

It was concluded that clinical test clusters to assess the presence or absence of discogenic pain as described by Laslett, Öberg, et al. (2005) and Laslett, Aprill, et al. (2006); radiculopathy of discogenic origin as described by Iverson et al. (2013); and

FJS as described by Laslett et al. (2004) and Laslett, McDonald, et al. (2006) were recommended for clinical examination of CLBP patients.

10.1.2 Objective 2

To report intra-examiner reliability of lumbar spine and neuro-dynamic flexibility measurements in a simulated CLBP population.

High intra-examiner reliability was established for measurement of lumbar spine ROM and neuro-dynamic flexibility in an overweight older population selected to mimic the anthropometry of a CLBP population. Ideally, the study would have been conducted on the population of interest, but there was an ethical dilemma and threat to reliability by exacerbation of symptoms in individuals with CLBP. The high reliability established in this simulated study population supported inclusion of these measurements in the clinical examination of CLBP.

10.1.3 Objective 3

On the basis of the outcomes of objectives 1 and 2, establish and apply a physiotherapy CLBP movement and patho-anatomical-based examination algorithm, along with age, gender, BMI, RM and ODI disability scores, VAS pain intensity, pain distribution, and MSPQ pain somatisation scores, to report the classification characteristics of a CLBP population.

This was the first research to classify and report the classification characteristics of a CLBP population using the MK-C. At first attendance, the MK-C provided final diagnostic classification in 94% of the CLBP population. The characteristics of the CLBP population were:

1. Female subjects represented 65% of the CLBP population.
2. Male subjects reported significantly lower MSPQ pain somatisation and disability than did female subjects.

3. Chronic FJS was the most common classification in CLBP.

4. Chronic SIJS, SI and postural syndrome were classified in a minority of cases.

5. MSPQ pain somatisation rated 'distressed' across all MK-C classification categories, indicating elevated somatic awareness in CLBP.

6. There were no significant differences in ODI, BMI, VAS and symptom duration between any MK-C classification categories.

7. Participants classified as FJS were significantly older but had significantly lower MSPQ pain somatisation and RM disability scores than those classified as discogenic CLBP.

8. Age, disability and MSPQ pain somatisation appear to be distinguishing characteristics of CLBP.

9. The proportions of MK-C derangement and dysfunction classification diverged from those reported in studies where classification was based solely on a movement-based McKenzie assessment, and in heterogeneous or acute LBP populations.

Future research is required to report and compare whether the characteristics of CLBP patients from primary-care settings and multiple study sites extrapolate to those of the CLBP participants in this secondary-care population. In addition, research is needed to explore the role of modified somatic pain perception levels in CLBP because scores indicating 'distress' on the MSPQ were a characteristic across all the MK-C classification categories in this population.

10.1.4 Objective 4

In a CLBP population, determine the diagnostic agreement between the

clinical classification by an experienced physiotherapist, using a

comprehensive CLBP examination algorithm, and a reference standard,

comprising a CLBP specialised neurosurgeon's diagnosis derived from a

physical examination, MRI and DSAI, as clinically indicated. In addition,

report the combined clinical diagnostic agreement of both examiners relative

to the outcome of a reference standard DSAI to diagnose FJS and FJS-

related nerve root compromise.

This study indicated 'fair' after-chance diagnostic agreement between a physiotherapist conducting the MK-C clinical examination and available reference standards in a CLBP population. Where FJS or FJS with NR compromise was considered the primary diagnosis, a reference standard DSAI was conducted for confirmation. Combined physiotherapist and neurosurgeon diagnostic classification relative to the DSAI was truly positive in almost half of cases, and falsely positive in only 18% of cases.

The MK-C was conducted by the physiotherapist without access to medical notes and MRI results in this research. This reflected current practice where physiotherapists are unable to request advanced imaging, and often have no access to imaging reports. The 'fair' after-chance diagnostic classification agreement with reference standards reported in this study is sub-optimal. However, this was the first study to report level of diagnostic classification agreement using the MK-C. If available prior to the physiotherapy CLBP examination, imaging results might have positively influenced diagnostic classification and ultimately improved diagnostic agreement. Further research is required to investigate if increased interdisciplinary communication and information sharing, when all examiners have access to advanced diagnostic imaging results, results in improved diagnostic agreement and optimised CLBP management.

186

It is recommended that use of the MK-C in clinical practice for CLBP patients should be conducted with caution, and it should be considered an interim architype. Research is required to refine structural components and flow of the MK-C process is required, and to establish intra- and inter-examiner reliability of the MK-C and criterion-related concurrent classification validity, compared with additional reference standards, which were not available and therefore not featured in this research. This will provide evidence to establish clinical utility of the MK-C and its reproducibility to classify CLBP patients.

10.1.5 Objective 5

To trial the utility and intra-examiner reliability of a novel RTUI TrA image acquisition tool that standardises probe force, inclination and roll.

Use of a novel image acquisition tool (FPD) to conduct TrA RTUI demonstrated high intra-examiner reliability, superior to the poor to moderate reliability of free-hand RTUI previously reported in CLBP populations. Use of an FPD, which provided real-time feedback and consistent probe force and orientation during TrA RTUI capture, was a distinguishing methodological feature of this study. This provides the opportunity to standardise longitudinal measures and track changes in TrA-C over time. It is recommended that, in future research and clinical practice, standardisation of probe-to-skin force, inclination and roll should be applied to optimise RTUI TrA assessment.

Although this study reported results specific to a CLBP population, the findings suggest that an FPD has potential to optimise reliability of RTUI TrA assessment in a variety of clinical populations and contexts. Therefore, it is recommended that future FPD studies investigate intra- and inter-examiner reliability of an FPD across multiple examiners of various experiences, and establish whether

FPD technology has potential to improve measurement reliability of other musculoskeletal structures.

As yet, FPD technology has not been integrated into RTUI machines and, although the methods described in this research remain inaccessible to other researchers and clinicians, future development of a machine-integrated FPD is anticipated. Inter-institution and cross-discipline studies between James Cook University physiotherapy and Massachusetts Institute of Technology bioengineering are expected to establish the utility, commercial viability, validity and reliability of RTUI machine-integrated FPD technology.

10.1.6 Objective 6

To assess TrA-C using reliable RTUI methodology before and immediately after pain abolition achieved from the anaesthetising effect of DSAI in CLBP.

In a CLBP population, the abolition of pain did not automatically improve TrA function. The rigorous, standardised, and highly reliable RTUI used in this study was subject to minimal measurement bias, and results supported the hypothesis that pain inhibition is not solely responsible for deficient TrA activation in CLBP. From a clinical perspective, CLBP relief as a standalone treatment is sub-optimal to immediately generate TrA recovery.

This study suggests that physiotherapy-specific TrA-C retraining for CLBP patients is required immediately following pain-relieving interventions, and should be considered in routine multidisciplinary management strategies for CLBP. Further research is required to determine optimal retraining and facilitation methods for TrA function; to determine if, following a pain-relieving injection, a 'refractory' TrA activation period exists, as well as its duration; and to determine if immediate post-injection TrA retraining increases the longevity of pain relief over and above that

which is currently achieved by DSAI alone. Further, CLBP avoidance behaviour was a potential confounding factor in this study, as its contribution to perpetual TrA-C inhibition is unclear. Therefore, it is recommended that future research investigate the effect of CLBP avoidance behaviour on TrA-C following pain relief.

10.2 Concluding Statement

This research has made a unique contribution to knowledge on diagnostic classification, physical assessment and RTUI in CLBP. No similar studies have been conducted where a physiotherapy examination algorithm comparable to the MK-C has been presented, trialled and applied to classify CLBP. Additionally, this is the first time measurement of TrA using an FPD, to standardise probe force, inclination and roll during RTUI, has been reported.

The individual studies have:

1. highlighted a critical need for research specific to physiotherapy examination and classification in CLBP

2. provided research evidence related to these needs by presenting the MK-C, a physiotherapy CLBP examination algorithm, which includes valid examination processes and measurements of lumbar spine ROM with reported reliability. (The MK-C has been trialled and applied in two studies, and can be adopted for future interventional research.)

3. described characteristics of a CLBP population examined with the MK-C and reported differences in these characteristics between classification categories, which provides knowledge to guide clinical reasoning

4. demonstrated 'Fair' diagnostic agreement between a physiotherapist using the MK-C and available reference standards

5. reported a 46% true-positive combined physiotherapist and specialist neurosurgeon's diagnostic classification, compared with a DSAI reference standard for FJS or FJS with NR compromise

6. indicated that, although not optimal, the level of MK-C diagnostic agreement provides clinicians with knowledge of its value and suitability as an evidence-based architype examination algorithm to classify CLBP. (However, further research is required to investigate the algorithm's performance to determine if the outcomes of this research may be generalised to other CLBP populations, and to guide future MK-C refinements.)

7. demonstrated utility and improved reliability for TrA RTUI using an FPD technology in CLBP, which can be used to improve measurement accuracy in future studies and in other musculoskeletal RTUI practice. (No such standardised method has been established previously for RTUI TrA measurement, or established such superior levels of reliability.)

8. provided research evidence of the complex relationship between pain abolition and TrA function in CLBP, indicating that resolution of pain does not result in any immediate change of TrA-C in CLBP. (This suggests that pain inhibition is not solely responsible for TrA dysfunction in CLBP, which, prior to this study, had been the subject of hypothetical assumption. This finding supports current theory that a complex relationship between persistent pain, neuromuscular control and pain-related muscle inhibition exists in CLBP. Importantly, it indicates that, following CLBP abolition, regardless of the mechanism by which this

occurs, physiotherapy-guided TrA retraining may provide improved outcomes, and should be a requisite component of patient management.)

Within professional scope of practice, and the biopsychosocial framework of CLBP management, the results and recommendations reported in this thesis have highlighted the critical need to identify and amalgamate reliable and valid CLBP examination processes, and have provided new knowledge about CLBP characteristics, classification and TrA function, with the ultimate aim to improve examination, classification and ultimately derive improved patients outcomes. In combination, the outcomes detailed and stated limitations of all included studies in this research highlight the complexity and challenges for research into physiotherapy examination of this LBP subgroup. Nevertheless, the outcomes of this research add knowledge on which to base future studies related to CLBP examination, physical assessment and RTUI methods, with potential to improve physiotherapy clinical practice and patient outcomes.

References

Abbott, J. H., Fritz, J. M., McCane, B., Shultz, B., Herbison, P., Lyons, B., …
Walsh, R. M. (2006). Lumbar segmental mobility disorders: Comparison of
two methods of defining abnormal displacement kinematics in a cohort of
patients with non-specific mechanical low back pain. *BMC Musculoskeletal
Disorders*, *7*, 45. doi:10.1186/1471-2474-7-45

Abbott, J. H., McCane, B., Herbison, P., Moginie, G., Chapple, C., & Hogarty, T.
(2005). Lumbar segmental instability: A criterion-related validity study of
manual therapy assessment. *BMC Musculoskeletal Disorders*, *6*.
doi:10.1186/1471-2474-6-56

Adhia, D. B., Bussey, M. D., Ribeiro, D. C., Tumilty, S., & Milosavljevic, S. (2013).
Validity and reliability of palpation-digitization for non-invasive kinematic
measurement—A systematic review. *Manual Therapy*, *18*, 26–34.
doi:10.1016/j.math.2012.06.004

Ahn, K., & Jhun, H.-J. (2015). New physical examination tests for lumbar
spondylolisthesis and instability: Low midline sill sign and interspinous gap
change during lumbar flexion-extension motion. *BMC Musculoskeletal
Disorders*, *16*, 97. doi:10.1186/s12891-015-0551-0

Al Nezari, N. H., Schneiders, A. G., & Hendrick, P. A. (2013). Neurological
examination of the peripheral nervous system to diagnose lumbar spinal disc
herniation with suspected radiculopathy: A systematic review and meta-
analysis. *The Spine Journal*, *13*, 657–674. doi:10.1016/j.spinee.2013.02.007

Ali, A., Khan, S. A., Aurangzeb, A., Ahmed, E., Ali, G., Muhammad, G., &
Mehmood, S. (2013). Lumbar disc herniation in patients with chronic
backache. *Journal of Ayub Medical College, Abbottabad*, *25*(3–4), 68–70.

Alonso, Y. (2004). The biopsychosocial model in medical research: The evolution of the health concept over the last two decades. *Patient Education and Counseling*, *53*, 239–244. doi:10.1016/S0738-3991(03)00146-0

Alqarni, A. M., Schneiders, A. G., & Hendrick, P. A. (2011). Clinical tests to diagnose lumbar segmental instability: A systematic review. *Journal of Orthopaedic & Sports Physical Therapy*, *41*, 130–140. doi:10.2519/jospt.2011.3457

Anderson Worth, S. G., Henry, S. M., & Bunn, J. Y. (2007). Real-time ultrasound feedback and abdominal hollowing exercises for people with low back pain. *New Zealand Journal of Physiotherapy*, *35*, 4–11.

Arab, A. M., Rasouli, O., Amiri, M., & Tahan, N. (2013). Reliability of ultrasound measurement of automatic activity of the abdominal muscle in participants with and without chronic low back pain. *Chiropractic & Manual Therapies*, *21*. doi:10.1186/2045-709X-21-37

Ardic, F., & Toraman, F. (2002). Psychological dimensions of pain in patients with rheumatoid arthritis, fibromyalgia syndrome, and chronic low back pain. *Journal of Musculoskeletal Pain*, *10*(4), 19–29.

Australian Institute of Health and Welfare. (2016). *Impacts of chronic back problems* (bulletin 137, cat. no. Aus 204). Retrieved from http://www.aihw.gov.au/publication-detail/?id=60129556196

Barker, P. J., Briggs, C. A., & Bogeski, G. (2004). Tensile transmission across the lumbar fasciae in unembalmed cadavers: Effects of tension to various muscular attachments. *Spine*, *29*, 129–138. doi:10.1097/01.BRS.0000107005.62513.32

Barker, P. J., Guggenheimer, K. T., Grkovic, I., Briggs, C. A., Jones, D., Thomas, L., & Hodges, P. W. (2006). Effects of tensioning the lumbar fasciae on segmental stiffness during flexion and extension: Young Investigator Award winner. *Spine*, *31*, 397–405. doi:10.1097/01.brs.0000195869.18844.56

Beattie, P., Rothstein, J. M., & Lamb, R. L. (1987). Reliability of the attraction method for measuring lumbar spine backward bending. *Physical Therapy*, *67*, 364–369.

Bertilson, B. C., Bring, J., Sjöblom, A., Sundell, K., & Strender, L.-E. (2006). Inter-examiner reliability in the assessment of low back pain (LBP) using the Kirkaldy-Willis classification (KWC). *European Spine Journal*, *15*, 1695–1703. doi:10.1007/s00586-005-0050-3

Boden, S. D., Davis, D. O., Dina, T. S., Patronas, N. J., & Wiesel, S. W. (1990). Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *Journal of Bone and Joint Surgery. American Volume*, *72*, 403.

Bogduk, N. (2004). Management of chronic low back pain. *Medical Journal of Australia, 180*(2), 79-83.

Boswell, M. V., Manchikanti, L., Kaye, A. D., Bakshi, S., Gharibo, C. G., Gupta, S., … Hirsch, J. A. (2015). A best-evidence systematic appraisal of the diagnostic accuracy and utility of facet (zygapophysial) joint injections in chronic spinal pain. *Pain Physician*, *18*(4), E497–533.

Brahee, D. D., Ogedegbe, C., Hassler, C., Nyirenda, T., Hazelwood, V., Morchel, H., … Feldman, J. (2013). Body mass index and abdominal ultrasound image quality: A pilot survey of sonographers. *Journal of Diagnostic Medical Sonography*, *29*, 66–72. doi:10.1177/8756479313476919

195

Broadbent, C. R., Maxwell, W. B., Ferrie, R., Wilson, D. J., Gawne-Cain, M., & Russell, R. (2000). Ability of anaesthetists to identify a marked lumbar interspace. *Anaesthesia*, *55*, 1122–1126. doi:10.1046/j.1365-2044.2000.01547-4.x

Brown, R. W., Cheng, Y. N., & Haacke, E. M.. (2014). *Magnetic resonance imaging: Physical principles and sequence design* [ebook version]. Retrieved from http://ebookcentral.proquest.com.elibrary.jcu.edu.au

Butler, D., Trafimow, J. H., Andersson, G. B. J., & McNeill, T. W. (1990). Discs degenerate before facets. *Spine*, *15*, 111–113.

Cameron, D. M., Bohannon, R. W., & Owen, S. V. (1994). Influence of hip position on measurements of the straight leg raise test. *Journal of Orthopaedic & Sports Physical Therapy*, *19*, 168–172.

Campbell, P. G., & Szmukler, G. I. (1993). The social state. *Psychiatric Bulletin*, *17*, 4–7.

Carlsson, H., & Rasmussen-Barr, E. (2013). Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. *Manual Therapy*, *18*, 103–110. doi:10.1016/j.math.2012.08.004

Cedraschi, C., Nordin, M., Nachemson, A. L., & Vischer, T. L. (1998). Health care providers should use a common language in relation to low back pain patients. *Bailliere's Clinical Rheumatology*, *12*, 1–15. doi:10.1016/s0950-3579(98)80003-4

Chambers, H. (2013). Physiotherapy and lumbar facet joint injections as a combination treatment for chronic low back pain. A narrative review of lumbar facet joint injections, lumbar spinal mobilizations, soft tissue massage

and lower back mobility exercises. *Musculoskeletal Care*, *11*, 106–120. doi:10.1002/msc.1045

Charlin, B., Lubarsky, S., Millette, B., Crevier, F., Audétat, M.-C., Charbonneau, A., … Bourdy, C. (2012). Clinical reasoning processes: Unravelling complexity through graphical representation. *Medical Education*, *46*, 454–463. doi:10.1111/j.1365-2923.2012.04242.x

Chipchase, L. S., Thoirs, K., & Jedrzejczak, A. (2009). The effectiveness of real time ultrasound as a biofeedback tool for muscle retraining. *Physical Therapy Reviews*, *14*(2). doi:10.1179/174328809X405991

Chou, R., Qaseem, A., Snow, V., Casey, D., Cross, J. T., Shekelle, P., & Owens, D. K. (2007). Diagnosis and treatment of low back pain: A joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Annals of Internal Medicine*, *147*(7), 478.

Christe, G., Redhead, L., Legrand, T., Jolles, B. M., & Favre, J. (2016). Multi-segment analysis of spinal kinematics during sit-to-stand in patients with chronic low back pain. *Journal of Biomechanics*, *49*(10), 2060–2067. doi:10.1016/j.jbiomech.2016.05.015

Cimolin, V., Vismara, L., Galli, M., Zaina, F., Negrini, S., & Capodaglio, P. (2011). Effects of obesity and chronic low back pain on gait. *Journal of NeuroEngineering and Rehabilitation*, *8*(55). doi:10.1186/1743-0003-8-55

Clare, H. A., Adams, R., & Maher, C. G. (2004). Reliability of the McKenzie spinal pain classification using patient assessment forms. *Physiotherapy*, *90*, 114–119. doi:10.1016/j.physio.2004.05.001

Clare, H. A., Adams, R., & Maher, C. G. (2005). Reliability of McKenzie classification of patients with cervical or lumbar pain. *Journal of*

Manipulative and Physiological Therapeutics, *28*, 122–127.

doi:10.1016/j.jmpt.2005.01.003

Clarkson, H. M. (2005). *Joint motion and function assessment: A research based practical guide*. Philadelphia, PA: Lippincott Williams and Wilkins.

Costa, L. O. P., Maher, C. G., Latimer, J., Hodges, P. W., & Shirley, D. (2009). An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain. *European Spine Journal*, *18*, 1059–1065. doi:10.1007/s00586-009-1018-5

Costa, L. O. P., Maher, C. G., Latimer, J., & Smeets, R. J. E. M. (2009). Reproducibility of rehabilitative ultrasound imaging for the measurement of abdominal muscle activity: A systematic review. *Physical Therapy*, *89*, 756–769. doi:10.2522/ptj.20080331

Critchley, D. J., & Coutts, F. J. (2002). Abdominal muscle function in chronic low back pain patients: Measurement with real-time ultrasound scanning. *Physiotherapy*, *88*, 322–332.

Curatolo, M., & Bogduk, N. (2010). Diagnostic blocks for chronic pain. *Scandinavian Journal of Pain*, *1*(4), 186–192. doi:10.1016/j.sjpain.2010.07.001

Dahl, A. K., Fauth, E. B., Ernsth-Bravell, M., & Hassing, L. B. (2013). Body mass index, change in body mass index, and survival in old and very old persons. *Journal of the American Geriatrics Society*, *61*, 512–518. doi:10.1111/jgs.12158

Dankaerts, W., O'Sullivan, P. B., Straker, L. M., Burnett, A. F., & Skouen, J. S. (2006). The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment.

Manual Therapy, *11*, 28–39.

doi:http://dx.doi.org/10.1016/j.math.2005.02.001

Datta, S., Everett, C. R., Trescot, A. M., Schultz, D. M., Adlaka, R., Abdi, S., …
Shah, R. V. (2007). An updated systematic review of the diagnostic utility of
selective nerve root blocks. *Pain Physician*, *10*(1), 113–128.

DePalma, M. J., Ketchum, J. M., & Saullo, T. (2011). What is the source of chronic
low back pain and does age play a role? *Pain Medicine*, *12*, 224–233.

DePalma, M. J., Ketchum, J. M., & Saullo, T. R. (2012). Multivariable analyses of
the relationships between age, gender, and body mass index and the source of
chronic low back pain. *Pain Medicine, 13*(4), 498-506. doi:10.1111/j.1526-
4637.2012.01339.x

Deyo, R. A., Dworkin, S. F., Amtmann, D., Andersson, G., Borenstein, D., Carragee,
E., … Weiner, D. K. (2014). Report of the NIH Task Force on research
standards for chronic low back pain. *The Spine Journal*, *14*(8), 1375–1391.
doi:10.1016/j.spinee.2014.05.002

Doriot, N., & Wang, X. (2006). Effects of age and gender on maximum voluntary
range of motion of the upper body joints. *Ergonomics*, *49*, 269–281.
doi:10.1080/00140130500489873

Dubois, J., Abboud, J., St-Pierre, C., Piché, M., & Descarreaux, M. (2014).
Neuromuscular adaptations predict functional disability independently of
clinical pain and psychological factors in patients with chronic non-specific
low back pain. *Journal of Electromyography and Kinesiology*, *24*, 550–557.
doi:10.1016/j.jelekin.2014.04.012

Dunn, K. M., & Croft, P. R. (2006). The importance of symptom duration in

    determining prognosis. *Pain (Amsterdam)*, *121*, 126–132.

    doi:10.1016/j.pain.2005.12.012

Eirikstoft, H., & Kongsted, A. (2014). Patient characteristics in low back pain

    subgroups based on an existing classification system. A descriptive cohort

    study in chiropractic practice. *Manual Therapy*, *19*(1), 65–71.

    doi:10.1016/j.math.2013.07.007

Engel, G. L. (1981). The clinical application of the biopsychosocial model. *Journal*

    of Medicine and Philosophy, *6*, 101.

Fairbank, J., Gwilym, S. E., France, J. C., Daffner, S. D., Dettori, J., Hermsmeyer, J.,

    & Andersson, G. (2011). The role of classification of chronic low back pain.

    Spine, *36*(Suppl. 21), S19–42. doi:10.1097/BRS.0b013e31822ef72c

Fairbank, J. C. T., Couper, J., Davies, J. B., & O'Brien, J. P. (1980). The Oswestry

    low back pain questionnaire. *Physiotherapy*, *66*, 271–273.

Fairbank, J. C. T., & Pynsent, P. B. (2000). The Oswestry Disability Index. *Spine*,

    25, 2940–2953. doi:10.1097/00007632-200011150-00017

Ferguson, S. A., Gallagher, S., & Marras, W. S. (2003). Validity and reliability of

    sincerity test for dynamic trunk motions. *Disability and Rehabilitation*, *25*,

    236–241. doi:10.1080/0963828021000030945

Ferreira, P. H., Ferreira, M. L., & Hodges, P. W. (2004). Changes in recruitment of

    the abdominal muscles in people with low back pain: Ultrasound

    measurement of muscle activity. *Spine*, *29*, 2560–2566.

    doi:10.1097/01.brs.0000144410.89182.f9

Ferreira, P. H., Ferreira, M. L., Maher, C. G., Refshauge, K. M., Latimer, J., &

    Adams, R. D. (2013). The therapeutic alliance between clinicians and patients

predicts outcome in chronic low back pain. *Physical Therapy*, *93*, 470–478. doi:10.2522/ptj.20120137

Flavell, C. A., Gordon, S., & Marshman, L. (2016). Classification characteristics of a chronic low back pain population using a combined McKenzie and patho-anatomical assessment. *Manual Therapy*, *26*, 201–207. doi:10.1016/j.math.2016.10.002

Flavell, C., Gordon, S., & Marshman, L. (2017a). *Measurement of transversus abdominis activation in chronic low back pain patients using a novel standardised real-time ultrasound imaging method.* Manuscript submitted for publication.

Flavell, C., Gordon, S., & Marshman, L. (2017b). Reply to the letter to the editor regarding 'Classification characteristics of a chronic low back pain population using a combined McKenzie and patho-anatomical assessment'. *Musculoskeletal Science and Practice*, *27*, e7. doi:10.1016/j.msksp.2016.12.012

Flavell, C., Gordon, S., Marshman, L., & Watt, K. (2014). Inter-rater reliability of classification systems in chronic low back pain populations. *Physical Therapy Reviews, 19*(3), 204-212. doi:10.1179/1743288X13Y.0000000131

Flavell, C. A., Gordon, S., & Watt, K. (2017). Intra-examiner reliability of lumbar spine and neuro-dynamic flexibility measurements in an older and overweight healthy asymptomatic population. *Journal of Back and Musculoskeletal Rehabilitation*, *30*, 79–84. doi:10.3233/BMR-160717

Ford, J., Story, I., O'Sullivan, P., & McMeeken, J. (2007). Classification systems for low back pain: A review of the methodology for development and validation. *Physical Therapy Reviews*, *12*, 33–42. doi:10.1179/108331907X174961

Ford, J. J., Hahne, A. J., Surkitt, L. D., Chan, A. Y. P., Richards, M. C., Slater, S. L.,…Taylor, N. F. (2016). Individualised physiotherapy as an adjunct to guideline-based advice for low back disorders in primary care: a randomised controlled trial. *British Journal of Sports Medicine, 50*(4), 237. doi:10.1136/bjsports-2015-095058

Freburger, J. K., Holmes, G. M., Agans, R. P., Jackman, A. M., Darter, J. D., Wallace, A. S., … Carey, T. S. (2009). The rising prevalence of chronic low back pain. *Archives of Internal Medicine (1960)*, *169*(3), 251–258. doi:10.1001/archinternmed.2008.543

Fritz, J. M., Beneciuk, J. M., George, S. Z., Hill, J. C., & Hay, E. M. (2011). Relationship between categorization with the STarT Back screening tool and prognosis for people receiving physical therapy for low back pain/invited commentary/author response. *Physical Therapy*, *91*, 722–736. doi:10.2522/ptj.20100109

Fritz, J. M., Brennan, G. P., Clifford, S. N., Hunter, S. J., & Thackeray, A. (2006). An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. *Spine*, *31*, 77–82. doi:10.1097/01.brs.0000193898.14803.8a

Fritz, J. M., Piva, S. R., & Childs, J. D. (2005). Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. *European Spine Journal*, *14*, 743–750. doi:10.1007/s00586-004-0803-4

Frost, M., Stuckey, S., Smalley, L. A., & Dorman, G. (1982). Reliability of measuring trunk motions in centimeters. *Physical Therapy*, *62*, 1431–1437.

Fujiwara, A., Tamai, K., Yamato, M., An, H. S., Yoshida, H., Saotome, K., & Kurihashi, A. (1999). The relationship between facet joint osteoarthritis and

disc degeneration of the lumbar spine: An MRI study. *European Spine Journal*, *8*, 396–401. doi:10.1007/s005860050193

Gilbertson, M. W., & Anthony, B. W. (2013). An ergonomic, instrumented ultrasound probe for 6-axis force/torque measurement. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 140–143). Piscataway, NJ: IEEE Service Center. doi:10.1109/EMBC.2013.6609457

Gill, K., Krag, M. H., Johnson, G. B., Haugh, L. D., & Pope, M. H. (1988). Repeatability of four clinical methods for assessment of lumbar spinal motion. *Spine*, *13*, 50–53.

Gnat, R., Saulicz, E., & Miądowicz, B. (2012). Reliability of real-time ultrasound measurement of transversus abdominis thickness in healthy trained subjects. *European Spine Journal*, *21*, 1508–1515. doi:10.1007/s00586-012-2184-4

Goddard, M. D., & Reid, J. D. (1965). Movements induced by straight leg raising in the lumbo-sacral roots, nerves and plexus, and in the intrapelvic section of the sciatic nerve. *Journal of Neurology, Neurosurgery & Psychiatry*, *28*, 12–18.

Guclu, D. G., Guclu, O., Ozaner, A., Senormanci, O., & Konkan, R. (2012). The relationship between disability, quality of life and fear-avoidance beliefs in patients with chronic low back pain. *Turkish Neurosurgery*, *22*(6), 724–731. doi:10.5137/1019-5149.JTN.6156-12.1

Hall, H., McIntosh, G., & Boyle, C. (2009). Effectiveness of a low back pain classification system. *The Spine Journal*, *9*, 648–657. doi:10.1016/j.spinee.2009.04.017

Hancock, M., Maher, C., Macaskill, P., Latimer, J., Kos, W., & Pik, J. (2012). MRI findings are more common in selected patients with acute low back pain than

controls? *European Spine Journal*, *21*, 240–246. doi:10.1007/s00586-011-1955-7

Hancock, M. J., Maher, C. G., Latimer, J., Spindler, M. F., McAuley, J. H., Laslett, M., & Bogduk, N. (2007). Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. *European Spine Journal*, *16*, 1539–1550. doi:10.1007/s00586-007-0391-1

Harris-Hayes, M., & Van Dillen, L. R. (2009). The inter-tester reliability of physical therapists classifying low back pain problems based on the Movement System Impairment classification system. *PM&R*, *1*, 117–126. doi:10.1016/j.pmrj.2008.08.001

Hawker, G. A., Mian, S., Kendzerska, T., & French, M. (2011). Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care & Research*, *63*(Suppl. 11), S240–S252. doi:10.1002/acr.20543

Hebert, J. J., Koppenhaver, S. L., Parent, E. C., & Fritz, J. M. (2009). A systematic review of the reliability of rehabilitative ultrasound imaging for the quantitative assessment of the abdominal and lumbar trunk muscles. *Spine*, *34*, E848–E856. doi:10.1097/BRS.0b013e3181ae625c

Hefford, C. (2008). McKenzie classification of mechanical spinal pain: Profile of syndromes and directions of preference. *Manual Therapy*, *13*, 75–81.

Heiss, D. G., Fitch, D. S., Fritz, J. M., Sanchez, W. J., Roberts, K. E., & Buford, J. A. (2004). The interrater reliability among physical therapists newly trained in a

classification system for acute low back pain. *Journal of Orthopaedic &*
Sports Physical Therapy, *34*, 430–439. doi:10.2519/jospt.2004.1555

Hengeveld, E., & Banks, K. (Eds.) (2014). *Maitland's vertebral manipulation.*
Volume 1: Management of neuromusculoskeletal disorders (8th ed.).
Edinburgh, United Kingdom: Churchill Livingstone.

Henry, S. M., Fritz, J. M., Trombley, A. R., & Bunn, J. Y. (2012). Reliability of a
treatment-based classification system for subgrouping people with low back
pain. *Journal of Orthopaedic & Sports Physical Therapy*, *42*, 797–805.
doi:10.2519/jospt.2012.4078

Henry, S. M., Van Dillen, L. R., Trombley, A. R., Dee, J. M., & Bunn, J. Y. (2013).
Reliability of novice raters in using the movement system impairment
approach to classify people with low back pain. *Manual Therapy*, *18*, 35–40.
doi:10.1016/j.math.2012.06.008

Henry, S. M., & Westervelt, K. C. (2005). The use of real-time ultrasound feedback
in teaching abdominal hollowing exercises to healthy subjects. *Journal of*
Orthopaedic & Sports Physical Therapy, *35*, 338–345.
doi:10.2519/jospt.2005.1757

Henschke, N., Maher, C. G., Refshauge, K. M., Herbert, R. D., Cumming. R. G.,
Bleasel, J., … McAuley, J. H. (2006). Prognosis of acute low back pain:
Design of a prospective inception cohort study. *BMC Musculoskeletal*
Disorders, *7*(54). doi:10.1186/1471-2474-7-54

Heuch, I., Hagen, K., Heuch, I., Nygaard, Ø., & Zwart, J.-A. (2010). The impact of
body mass index on the prevalence of low back pain: The HUNT study.
Spine, *35*, 764–768. doi:10.1097/BRS.0b013e3181ba1531

Hides, J. A. (2006). Rehabilitative ultrasound imaging of the lumbar multifidus muscle: Biofeedback training, clinical and research applications. *Journal of Orthopaedic & Sports Physical Therapy*, *36*, A-12. doi:10.2519/jospt.2006.0301

Hides, J. A., Belavý, D. L., Cassar, L., Williams, M., Wilson, S. J., & Richardson, C. A. (2009). Altered response of the anterolateral abdominal muscles to simulated weight-bearing in subjects with low back pain. *European Spine Journal*, *18*, 410–418. doi:10.1007/s00586-008-0827-2

Hides, J. A., Miokovic, T., Belavý, D. L., Stanton, W. R., & Richardson, C. A. (2007). Ultrasound imaging assessment of abdominal muscle function during drawing-in of the abdominal wall: An intrarater reliability study. *Journal of Orthopaedic & Sports Physical Therapy*, *37*, 480–486. doi:10.2519/jospt.2007.2416

Hides, J. A., Richardson, C. A., & Jull, G. A. (1996). Multifidus muscle recovery is not automatic after resolution of acute, first-episode low back pain. *Spine*, *21*, 2763–2769. doi:10.1097/00007632-199612010-00011

Hides, J. A., Richardson, C. A., Jull, G. A., & Davies, S. (1995). Ultrasound imaging in rehabilitation. *Australian Journal of Physiotherapy, 41*(3), 187–193. doi:10.1016/S0004-9514(14)60429-3

Hides, J., Stanton, W., Dilani Mendis, M., & Sexton, M. (2011). The relationship of transversus abdominis and lumbar multifidus clinical muscle tests in patients with chronic low back pain. *Manual Therapy*, *16*, 573–577. doi:10.1016/j.math.2011.05.007

Hides, J., Wilson, S., Stanton, W., McMahon, S., Keto, H., McMahon, K., … Richardson, C. (2006). An MRI investigation into the function of the

transversus abdominis muscle during "drawing-in" of the abdominal wall. *Spine*, *31*, E175–E178. doi:10.1097/01.brs.0000202740.86338.df

Hides, J. A., Wong, I., Wilson, S. J., Belavý, D. L., & Richardson, C. A. (2007). Assessment of abdominal muscle function during a simulated unilateral weight-bearing task using ultrasound imaging. *Journal of Orthopaedic & Sports Physical Therapy*, *37*, 467–471. doi:10.2519/jospt.2007.2417

Hill, J. C., Dunn, K. M., Lewis, M., Mullis, R., Main, C. J., Foster, N. E., & Hay, E. M. (2008). A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Care & Research*, *59*(5), 632–641. doi:10.1002/art.23563

Hill, J. C., Dunn, K. M., Main, C. J., & Hay, E. M. (2010). Subgrouping low back pain: A comparison of the STarT Back tool with the Örebro Musculoskeletal Pain Screening Questionnaire. *European Journal of Pain*, *14*, 83–89. doi:10.1016/j.ejpain.2009.01.003

Hill, J. C., Vohora, K., Dunn, K. M., Main, C. J., & Hay, E. M. (2010). Comparing the STarT Back screening tool's subgroup allocation of individual patients with that of independent clinical experts. *Clinical Journal of Pain*, *26*, 783–787. doi:10.1097/AJP.0b013e3181f18aac

Hodges, P. W., & Moseley, G. L. (2003). Pain and motor control of the lumbopelvic region: Effect and possible mechanisms. *Journal of Electromyography and Kinesiology*, *13*, 361–370. doi:10.1016/S1050-6411(03)00042-7

Hodges, P. W., Moseley, G. L., Gabrielsson, A., & Gandevia, S. C. (2003). Experimental muscle pain changes feedforward postural responses of the trunk muscles. *Experimental Brain Research*, *151*, 262–271. doi:10.1007/s00221-003-1457-x

Hong , J., Reed , C., Novick , D., & Happich , M. (2013). Costs associated with
treatment of chronic low back pain. *Health Services Research*, *38*(1), 75–82.
doi:10.1097/BRS.0b013e318276450f

Hopkins, W. (2000). Quantitative research design. *Sportscience*, 4(1). Retrieved from
http://sportsci.org/jour/0001/wghdesign.html

Hoy, D., Bain, C., Williams, G., March, L., Brooks, P., Blyth, F., … Buchbinder, R.
(2012). A systematic review of the global prevalence of low back pain.
Arthritis & Rheumatism, *64*(6), 2028–2037. doi:10.1002/art.34347

Huskisson, E. C. (1974). Measurement of pain. *The Lancet*, *304*, 1127–1131.
doi:10.1016/S0140-6736(74)90884-8

Hyytiäinen, K., Salminen, J. J., Suvitie, T., Wickström, G., & Pentti, J. (1991).
Reproducibility of nine tests to measure spinal mobility and trunk muscle
strength. *Scandinavian Journal of Rehabilitation Medicine*, *23*, 3–10.

International Maitland Teacher's Association. (2010). A tribute to the life and work
of G.D. Maitland 1924–2010 by the International Maitland Teachers
Association. *Manual Therapy*, *15*(3), 300–301.
doi:10.1016/j.math.2010.02.004

Ishida, H., & Watanabe, S. (2012). Influence of inward pressure of the transducer on
lateral abdominal muscle thickness during ultrasound imaging. *Journal of
Orthopaedic & Sports Physical Therapy, 42*(9), 815–818.
doi:10.2519/jospt.2012.4064

Iversen, T., Solberg, T., Romner, B., Wilsgaard, T., Nygaard, O., Waterloo, K., …
Ingebrigtsen, T. (2013). Accuracy of physical examination for chronic lumbar
radiculopathy. *BMC Musculoskeletal Disorders*, *14*, 206.

Jensen, M. C., Brant-Zawadzki, M. N., Obuchowski, N., Modic, M. T., Malkasian, D., & Ross, J. S. (1994). Magnetic resonance imaging of the lumbar spine in people without back pain. *New England Journal of Medicine*, *331*(2), 69–73. doi:10.1056/nejm199407143310201

Jones, G. T., Watson, K. D., Silman, A. J., Symmons, D. P. M., & Macfarlane Gary J. (2003). Predictors of low back pain in British schoolchildren: a population-based prospective cohort study. *Pediatrics*, *111*(4), 822-828.

Jull, G., & Moore, A. (2012). Hands on, hands off? The swings in musculoskeletal physiotherapy practice. *Manual Therapy*, *17*, 199–200. doi:10.1016/j.math.2012.03.009

Kamper, S. J., Apeldoorn, A. T., Chiarotto, A., Smeets, R. J. E. M., Ostelo, R. W. J. G., Guzman, J., & van Tulder, M. W. (2015). Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. *British Medical Journal (Online)*, *350*, 1–11. doi:10.1136/bmj.h444

Karayannis, N. V., Jull, G. A., & Hodges, P. W. (2012). Physiotherapy movement based classification approaches to low back pain: Comparison of subgroups through review and developer/expert survey. *BMC Musculoskeletal Disorders*, *13*, 24. doi:10.1186/1471-2474-13-24

Kennedy, D. J., Engel, A., Kreiner, D. S., Nampiaparampil, D., Duszynski, B., & MacVicar, J. (2015). Fluoroscopically guided diagnostic and therapeutic intra-articular sacroiliac joint injections: A systematic review. *Pain Medicine*, *16*(8), 1500–1518. doi:10.1111/pme.12833

Kiesel, K. B., Uhl, T., Underwood, F. B., & Nitz, A. J. (2008). Rehabilitative

    ultrasound measurement of select trunk muscle activation during induced

    pain. *Manual Therapy*, *13*, 132–138. doi:10.1016/j.math.2006.10.003

Kilpikoski, S., Airaksinen, O., Kankaanpää, M., Leminen, P., Videman, T., & Alen,

    M. (2002). Interexaminer reliability of low back pain assessment using the

    McKenzie method. *Spine*, *27*, E207–214.

Kim, K. H., Cho, S. H., Goo, B. O., & Baek, I. H. (2013). Differences in transversus

    abdominis muscle function between chronic low back pain patients and

    healthy subjects at maximum expiration: Measurement with real-time

    ultrasonography. *Journal of Physical Therapy Science*, *25*, 861–863.

    doi:10.1589/jpts.25.861

Kongsted, A., & Leboeuf-Yde, C. (2010). The Nordic back pain subpopulation

    program: Can low back pain patterns be predicted from the first consultation

    with a chiropractor? A longitudinal pilot study. *Chiropractic & Osteopathy*,

    18(1), 8. doi:10.1186/1746-1340-18-8

Koppenhaver, S. L., Hebert, J. J., Fritz, J. M., Parent, E. C., Teyhen, D. S., & Magel,

    J. S. (2009). Reliability of rehabilitative ultrasound imaging of the transversus

    abdominis and lumbar multifidus muscles. *Archives of Physical Medicine &*

    Rehabilitation, *90*(1), 87–94. doi:10.1016/j.apmr.2008.06.022

Koppenhaver, S. L., Hebert, J. J., Parent, E. C., & Fritz, J. M. (2009). Rehabilitative

    ultrasound imaging is a valid measure of trunk muscle size and activation

    during most isometric sub-maximal contractions: a systematic review.

    Australian Journal of Physiotherapy, *55*, 153–169. doi:10.1016/S0004-

    9514(09)70076-5

Koppenhaver, S. L., Parent, E. C., Teyhen, D. S., Hebert, J. J., & Fritz, J. M. (2009). The effect of averaging multiple trials on measurement error during ultrasound imaging of transversus abdominis and lumbar multifidus muscles in individuals with low back pain. *Journal of Orthopaedic & Sports Physical Therapy*, *39*, 604–611. doi:10.2519/jospt.2009.3088

Kotilainen, E., & Valtonen, S. (1993). Clinical instability of the lumbar spine after microdiscectomy. *Acta Neurochirurgica (Wein)*, *125*, 120–126.

Krag, M. H., Byrne, K. B., Haugh, L. D., Miller, L., Johnson, G. B., & Pope, M. H. (1987). Objective determination of trunk muscle dimensions using ultrasound. *IEEE Engineering in Medicine and Biology Magazine: The Quarterly Magazine of the Engineering in Medicine & Biology Society*, *6*(3), 27–30. doi:10.1109/MEMB.1987.5006434

Kuo, Y.-L., Tully, E. A., & Galea, M. P. (2008). Skin movement errors in measurement of sagittal lumbar and hip angles in young and elderly subjects. *Gait & Posture*, *27*, 264–270. doi:10.1016/j.gaitpost.2007.03.016

Kuslich, S. D., Ulstrom, C. L., & Michael, C. J. (1991). The tissue origin of low back pain and sciatica: A report of pain response to tissue stimulation during operations on the lumbar spine using local anesthesia. *Orthopedic Clinics of North America*, *22*(2), 181.

Kutschenko, L. K. (2011). In quest of 'good' medical classification systems. *Medicine Studies*, *3*, 53–70. doi:10.1007/s12376-011-0065-5

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. doi:10.2307/2529310

Lariviere, C., Gagnon, D., De Oliveira, E., Jr., Henry, S. M., Mecheri, H., & Dumas, J. P. (2013). Reliability of ultrasound measures of the transversus abdominis:

Effect of task and transducer position. *PM&R*, *5*, 104–113.

doi:10.1016/j.pmrj.2012.11.002

Laslett, M. (2008). Evidence-based diagnosis and treatment of the painful sacroiliac

joint. *Journal of Manual & Manipulative Therapy*, *16*, 142–152.

doi:10.1179/jmt.2008.16.3.142

Laslett, M., Aprill, C. N., McDonald, B., & Öberg, B. (2006). Clinical predictors of

lumbar provocation discography: A study of clinical predictors of lumbar

provocation discography. *European Spine Journal*, *15*, 1473–1484.

doi:10.1007/s00586-006-0062-7

Laslett, M., Aprill, C. N., McDonald, B., & Young, S. B. (2005). Diagnosis of

sacroiliac joint pain: Validity of individual provocation tests and composites

of tests. *Manual Therapy*, *10*, 207–218. doi: 10.1016/j.math.2005.01.003

Laslett, M., McDonald, B., Aprill, C. N., Tropp, H., & Öberg, B. (2006). Clinical

predictors of screening lumbar zygapophyseal joint blocks: Development of

clinical prediction rules. *The Spine Journal*, *6*, 370–379.

doi:10.1016/j.spinee.2006.01.004

Laslett, M., McDonald, B., Tropp, H., Aprill, C. N., & Öberg, B. (2005). Agreement

between diagnoses reached by clinical examination and available reference

standards: A prospective study of 216 patients with lumbopelvic pain. *BMC*

Musculoskeletal Disorders, *6*, 28. doi:10.1186/1471-2474-6-28

Laslett, M., Öberg, B., Aprill, C. N., & McDonald, B. (2004). Zygapophysial joint

blocks in chronic low back pain: A test of Revel's model as a screening test.

BMC Musculoskeletal Disorders, *5*, 1–6. doi:10.1186/1471-2474-5-43

Laslett, M., Öberg, B., Aprill, C. N., & McDonald, B. (2005). Centralization as a

predictor of provocation discography results in chronic low back pain, and

the influence of disability and distress on diagnostic power. *The Spine Journal*, *5*, 370–380. doi:10.1016/j.spinee.2004.11.007

Laslett, M., & van Wijmen, P. (1999). Low back and referred pain: Diagnosis and a proposed new system of classification. *New Zealand Journal of Physiotherapy*, *27*(2), 5–14.

Laslett, M., Young, S. B., Aprill, C. N., & McDonald, B. (2003). Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests. *Australian Journal of Physiotherapy*, *49*, 89–97.

Laupacis, A., Sekar, N., & Stiell, I. G. (1997). Clinical prediction rules: A review and suggested modifications of methodological standards. Journal of the American Medical Association, 277(6), 488-494.

Leard, J. S., Crane, B. A., & Ball, K. A. (2009). Intrarater and interrater reliability of 22 clinical measures associated with lower quarter malalignment. *Journal of Manipulative and Physiological Therapeutics*, *32*, 270–276. doi:10.1016/j.jmpt.2009.03.009

Leclaire, R., Blier, F., Fortin, L., & Proulx, R. (1997). A cross-sectional study comparing the Oswestry and Roland–Morris functional disability scales in two populations of patients with low back pain of different levels of severity. *Spine*, *22*, 68–71. doi:10.1097/00007632-199701010-00011

Lieu, D. (2010). Ultrasound physics and instrumentation for pathologists. *Archives of Pathology & Laboratory Medicine, 134*(10), 1541–1556.

Lipschik, G. Y., Von Feldt, J. M., & Frame, L. (2009). *Oxford American handbook of clinical diagnosis*. Retrieved from http://ebookcentral.proquest.com.elibrary.jcu.edu.au/lib/jcu/detail.action?docID=2033532

Littlewood, C., & May, S. (2007). Measurement of range of movement in the lumbar

    spine—what methods are valid? A systematic review. *Physiotherapy*, *93*,

    201–211. doi:10.1016/j.physio.2006.10.006

Long, D. M., BenDebba, M., Torgerson, W. S., Boyd, R. J., Dawson, E. G., Hardy,

    R. W., … Watts, C. (1996). Persistent back pain and sciatica in the United

    States. *Journal of Spinal Disorders*, *9*(1), 40–58. doi:10.1097/00002517-

    199602000-00007

Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a

    quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal*

    of Clinical Epidemiology, *63*, 854–861. doi:10.1016/j.jclinepi.2009.10.002

Macrae, I. F., & Wright, V. (1969). Measurement of back movement. *Annals of the*

    Rheumatic Diseases, *28*, 584. doi:10.1136/ard.28.6.584

Main, C. J. (1983). The Modified Somatic Perception Questionnaire (MSPQ).

    Journal of Psychosomatic Research, *27*, 503–514. doi:10.1016/0022-

    3999(83)90040-5

Main, C. J., Sowden, G., Hill, J. C., Watson, P. J., & Hay, E. M. (2012). Integrating

    physical and psychological approaches to treatment in low back pain: The

    development and content of the STarT Back trial's 'high-risk' intervention.

    Physiotherapy, *98*, 110–116. doi:10.1016/j.physio.2011.03.003

Main, C. J., Wood, P. L. R., Hollis, S., & Spanswick, C. C. (1992). The Distress and

    Risk Assessment Method: A simple patient classification to identify distress

    and evaluate the risk of poor outcome. *Spine, 17*(1), 42–52.

    doi:10.1097/00007632-199201000-00007

Maitland, G. O. (1986). The slump test: Examination and treatment. *Australian*

    Journal of Physiotherapy, *31*, 215–219.

Mannion, A. F., Caporaso, F., Pulkovski, N., & Sprott, H. (2012). Spine stabilisation exercises in the treatment of chronic low back pain: A good clinical outcome is not associated with improved abdominal muscle function. *European Spine Journal*, *21*, 1301–1310. doi:10.1007/s00586-012-2155-9

Mannion, A. F., Pulkovski, N., Gubler, D., Gorelick, M., O'Riordan, D., Loupas, T., … Sprott, H. (2008). Muscle thickness changes during abdominal hollowing: An assessment of between-day measurement error in controls and patients with chronic low back pain. *European Spine Journal*, *17*, 494–501. doi:10.1007/s00586-008-0589-x

Marras, W. S., Ferguson, S. A., Gupta, P., Bose, S., Parnianpour, M., Kim, J. Y., & Crowell, R. R. (1999). The quantification of low back disorder using motion measures: Methodology and validation. *Spine*, *24*, 2091–2100. doi:10.1097/00007632-199910150-00005

Martin, D. J., Wells, I. T. P., & Goodwin, C. R. (2015). Physics of ultrasound. *Anaesthesia and Intensive Care Medicine*, *16*, 132–135. doi:10.1016/j.mpaic.2015.01.003

May, S. (2013). Obituary: Robin McKenzie 1931–2013. *Manual Therapy*, *18*(4), 272–273. doi:10.1016/j.math.2013.05.012

May, S., & Aina, A. (2012). Centralization and directional preference: A systematic review. *Manual Therapy*, *17*, 497–506. doi:10.1016/j.math.2012.05.003

McCarthy, C. J., Arnall, F. A., Strimpakos, N., Freemont, A., & Oldham, J. A. (2004). The biopsychosocial classification of non-specific low back pain: A systematic review. *Physical Therapy Reviews*, *9*, 17–30. doi:10.1179/108331904225003955

McCarthy, C. J., Roberts, C., Gittins, M., & Oldham, J. A. (2012). A process of

    subgroup identification in non-specific low back pain using a standard

    clinical examination and cluster analysis. *Physiotherapy Research*

    International, *17*, 92–100. doi:10.1002/pri.517

McCarthy, C. J., Rushton, A., Billis, V., Arnall, F., & Oldham, J. A. (2006).

    Development of a clinical examination in non-specific low back pain: A

    Delphi technique. *Journal of Rehabilitation Medicine*, *38*, 263–267.

    doi:10.1080/16501970600632768

McKenzie Institute International. (2005). *The McKenzie Institute lumbar spine*

    assessment. Retrieved from

    http://mckenziemdt.org/forms/2006MIIAssessmentFormLumbar.pdf.

McKenzie, R. A. (1981). *The lumbar spine: Mechanical diagnosis and therapy*.

    Waikanae, New Zealand: Spinal Publications.

McKenzie, R. A., & May, S. (2003). *The lumbar spine mechanical diagnosis &*

    therapy (2nd ed., Vol. 2). Waikanae, New Zealand: Spinal Publications.

McMeeken, J. M., Beith, I. D., Newham, D. J., Milligan, P., & Critchley, D. J.

    (2004). The relationship between EMG and change in thickness of

    transversus abdominis. *Clinical Biomechanics*, *19*, 337–342.

    doi:10.1016/j.clinbiomech.2004.01.007

Miller, J. A. A., Schmatz, C., & Schultz, A. B. (1988). Lumbar disc degeneration:

    Correlation with age, sex, and spine level in 600 autopsy specimens. *Spine*,

    13, 173–178.

Miller, J. C. (2005). Imaging and obese patients. *Radiology Rounds: A Newsletter for*

    Referring Physicians, *3*(7), 1–4. Retrieved from

    http://www.mghradrounds.org/index.php?src=gendocs&link=2005_july

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting

    items for systematic reviews and meta-analyses: The PRISMA statement.

    Annals of Internal Medicine, *151*, 264.

Moore, J. E. (2010). Chronic Low Back Pain and Psychosocial Issues. *Physical*

    Medicine and Rehabilitation Clinics of North America, *21*(4), 801-815. doi:

    10.1016/j.pmr.2010.06.005

Morris, T., Hee, S. W., Stallard, N., Underwood, M., & Patel, S. (2015). Can we

    convert between outcome measures of disability for chronic low back pain?

    Spine, *40*, 734–739. doi:10.1097/BRS.0000000000000866

Moseley, G. L., Nicholas, M. K., & Hodges, P. W. (2004). Does anticipation of back

    pain predispose to back trouble? *Brain, 127*(10), 2339-2347. doi:

    10.1093/brain/awh248

Moult, E., Ungi, T., Welch, M., Lu, J., McGraw, R. C., & Fichtinger, G. (2013).

    Ultrasound-guided facet joint injection training using Perk Tutor.

    International Journal of Computer Assisted Radiology and Surgery, *8*(5),

    831–836. doi:10.1007/s11548-012-0811-5

Murphy, D. R., Hurwitz, E. L., & Nelson, C. F. (2008). A diagnosis-based clinical

    decision rule for spinal pain part 2: Review of the literature. *Chiropractic &*

    Osteopathy, *16*(1), 7. doi:10.1186/1746-1340-16-7

National Health and Medical Research Council. (2009). *NHMRC levels of evidence*

    and grades for recommendations for guideline developers. Retrieved from

    https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_le

    vels_grades_evidence_120423.pdf.

Nelson, D. A., & Landau, W. M. (2001). Intraspinal steroids: History, efficacy,

    accidentality, and controversy with review of United States Food and Drug

Administration reports. *Journal of Neurology, Neurosurgery, and Psychiatry*, *70*(4), 433–443. doi:10.1136/jnnp.70.4.433

Ngan, T. T., Tuan, T. M., Son, L. H., Minh, N. H., & Dey, N. (2016). Decision making based on fuzzy aggregation operators for medical diagnosis from dental x-ray images. *Journal of Medical Systems*, *40*(12), 280. doi:10.1007/s10916-016-0634-y

Norton, K., & Olds, T. (2006). *Anthropometrica: A textbook of body measurement for sports and health courses*. New Delhi, India: CBS Publishers & Distributers PVt.Ltd.

O'Sullivan, P. B. (2000). Lumbar segmental 'instability': Clinical presentation and specific stabilizing exercise management. *Manual Therapy*, *5*(1), 2–12. doi:10.1054/math.1999.0213

O'Sullivan, P. (2005). Diagnosis and classification of chronic low back pain disorders: Maladaptive movement and motor control impairments as underlying mechanism. *Manual Therapy*, *10*(4), 242–255. doi:10.1016/j.math.2005.07.001

Ortiz, S. H. C., Chiu, T., & Fox, M. D. (2012). Ultrasound image enhancement: A review. *Biomedical Signal Processing and Control*, *7*, 412–428. doi:10.1016/j.bspc.2012.02.002

Ota, M., & Kaneoka, K. (2011). Differences in abdominal muscle thicknesses between chronic low back pain patients and healthy subjects. *Journal of Physical Therapy Science*, *23*, 855–858. doi:10.1589/jpts.23.855

Panjabi, M. M. (1992). The stabilizing system of the spine. Part I. Function, dysfunction, adaptation, and enhancement. *Journal of Spinal Disorders*, *5*, 383–389.

Paris, S. V. (2000). A history of manipulative therapy through the ages and up to the current controversy in the United States. *Journal of Manual & Manipulative Therapy*, *8*(2), 66–77. doi:10.1179/106698100790819555

Paskowski, I., Schneider, M., Stevans, J., Ventura, J. M., & Justice, B. D. (2011). A hospital-based standardized spine care pathway: Report of a multidisciplinary, evidence-based process. *Journal of Manipulative and Physiological Therapeutics*, *34*, 98–106. doi:10.1016/j.jmpt.2010.12.004

Paul, B., Leitner, C., Vacariu, G., Wick, F., Zehetmayer, S., Matzner, M., … Ebenbichler, G. (2008). Low-back pain assessment based on the brief ICF core sets: Diagnostic relevance of motor performance and psychological tests. *American Journal of Physical Medicine & Rehabilitation*, *87*, 452–460. doi:10.1097/PHM.0b013e318174e37a

Pengel, L. H. M., Herbert, R. D., Maher, C. G., & Refshauge, K. M. (2003). Acute low back pain: Systematic review of its prognosis. *British Medical Journal*, *327*(7410), 323–323. doi: 10.1136/bmj.327.7410.323

Perkin, H., Bond, E., Thompson, J., Woods, E., & Smith, C. (2003). Real time ultrasound: An objective measure of skeletal muscle. *Physical Therapy Reviews*, *8*, 99–108. doi:10.1179/108331903225002506

Petersen, T., Laslett, M., & Juhl, C. (2017). Clinical classification in low back pain: Best-evidence diagnostic rules based on systematic reviews. *BMC Musculoskeletal Disorders*, *18*, 188. doi:10.1186/s12891-017-1549-6

Petersen, T., Laslett, M., Thorsen, H., Manniche, C., Ekdahl, C., & Jacobsen, S. (2003). Diagnostic classification of non-specific low back pain. A new system integrating pathoanatomic and clinical categories. *Physiotherapy Theory and Practice*, *19*, 213–237. doi:10.1080/09593980390246760

Petersen, T., Olsen, S., Laslett, M., Thorsen, H., Manniche, C., Ekdahl, C., & Jacobsen, S. (2004). Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain. *Australian Journal of Physiotherapy*, *50*, 85–91. doi:10.1016/S0004-9514(14)60100-8

Petersen, T., Thorsen, H., Manniche, C., & Ekdahl, C. (1999). Classification of non-specific low back pain: A review of the literature on classifications systems relevant to physiotherapy. *Physical Therapy Reviews*, *4*, 265–281. doi:10.1179/ptr.1999.4.4.265

Petty, N. J. (2013). *Neuromusculoskeletal examination and assessment: A handbook for therapists* (4th ed.). Sydney, NSW: Churchill Livingstone/Elsevier.

Portek, I., Pearcy, M. J., Reader, G. P., & Mowat, A. G. (1983). Correlation between radiographic and clinical measurement of lumbar spine movement. *British Journal of Rheumatology*, *22*(4), 197–205.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Prins, M. R., van der Wurff, P., & Groen, G. J. (2013). Chronic low back pain patients with accompanying leg pain: The relationship between pain extent and pain intensity, disability and health status. *Journal of Back and Musculoskeletal Rehabilitation*, *26*, 55–61. doi:10.3233/BMR-2012-00350

Pulkovski, N., Mannion, A. F., Caporaso, F., Toma, V., Gubler, D., Helbling, D., & Sprott, H. (2012). Ultrasound assessment of transversus abdominis muscle contraction ratio during abdominal hollowing: A useful tool to distinguish between patients with chronic low back pain and healthy controls? *European Spine Journal*, *21*(Suppl. 6), S750–S759.

Raney, N. H., Teyhen, D. S., & Childs, J. D. (2007). Observed changes in lateral abdominal muscle thickness after spinal manipulation: A case series using rehabilitative ultrasound imaging. *Journal of Orthopaedic & Sports Physical Therapy*, *37*, 472–479. doi:10.2519/jospt.2007.2523

Rasband, W. S. (1997). *ImageJ.* Retrieved from http://imagej.nih.gov/ij/

Rasouli, O., Arab, A. M., Amiri, M., & Jaberzadeh, S. (2011). Ultrasound measurement of deep abdominal muscle activity in sitting positions with different stability levels in subjects with and without chronic low back pain. *Manual Therapy*, *16*, 388–393. doi:10.1016/j.math.2011.01.009

Razmjou, H., Kramer, J. F., & Yamada, R. (2000). Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. *Journal of Orthopaedic & Sports Physical Therapy*, *30*, 368–389.

Refshauge, K. M., & Gass, E. M. (2004). *Musculoskeletal physiotherapy: Clinical science and evidence-based practice* (2nd ed.). Oxford, United Kingdom: Butterworth-Heinemann.

Reneman, M. F., Kuijer, W., Brouwer, S., Schiphorst Preuper, H. R., Groothoff, J. W., Geertzen, J. H. B., & Dijkstra, P. U. (2006). Symptom increase following a functional capacity evaluation in patients with chronic low back pain: An explorative study of safety. *Journal of Occupational Rehabilitation*, *16*(2), 192–205. doi:10.1007/s10926-006-9023-6

Revel, M., Poiraudeau, S., Auleley, G. R., Payan, C., Denke, A., Nguyen, M., … Fermanian, J. (1998). Capacity of the clinical picture to characterize low back pain relieved by facet joint anesthesia: Proposed criteria to identify patients with painful facet joints. *Spine*, *23*, 1972–1976. doi:10.1097/00007632-199809150-00011

Richardson, C., Hodges, P. W., & Hides, J. (2004). *Therapeutic exercise for lumbopelvic stabilization: A motor control approach for the treatment and prevention of low back pain* (Vol. 2). Edinburgh, United Kingdom: Churchill Livingstone.

Richardson, C. A., & Jull, G. A. (1995). Muscle control-pain control. What exercises would you prescribe? *Manual Therapy*, *1*, 2–10. doi:10.1054/math.1995.0243

Riddle, D. L., & Rothstein, J. M. (1993). Intertester reliability of McKenzie's classifications of the syndrome types present in patients with low back pain. *Spine*, *18*, 1333–1344.

Roddey, T. S., Brizzolara, K. J., & Cook, K. F. (2008). Side-to-side differences in the transverse abdominus muscle measured by real time ultrasound in persons with and without chronic low back pain. *Orthopaedic Physical Therapy Practice*, *20*(2), 56–59.

Roland, M. O., & Fairbank, J. C. T. (2000). The Roland–Morris disability questionnaire and the Oswestry Disability questionnaire. *Spine*, *25*, 3115–3124.

Roland, M. O., & Morris, R. W. (1983). A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low back pain. *Spine*, *8*, 141–144.

Rostami, M., Noormohammadpour, P., Mansournia, M. A., Hantoushzadeh, S., Farahbakhsh, F., Nourian, R., & Kordi, R. (2015). Comparison of the thickness of lateral abdominal muscles between pregnant women with and without low back pain. *PM&R*, *7*, 474–478. doi:10.1016/j.pmrj.2014.10.014

Rutjes, A. W. S., Reitsma, J. B., Coomarasamy, A., Khan, K. S., & Bossuyt, P. M. M. (2007). Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment*, *11*(50), iii–47.

Sahrmann, S. A., Norton, B. J., Caldwell, C. A., Bloom, N. J., McDonnell, M. K., & Van Dillen, L. R. (2003). Movement system impairment-based categories for low back pain: Stage I validation. *Journal of Orthopaedic & Sports Physical Therapy*, *33*, 126–142.

Saliba, S., Croy, T., Guthrie, R., Grooms, D., Weltman, A., & Grindstaff, T. (2010). Differences in transverse abdominis activation with stable and unstable bridging exercises in individuals with low back pain. *North American Journal of Sports Physical Therapy*, *5*(2).

Saragiotto, B. T., Maher, C. G., Hancock, M. J., & Koes, B. W. (2017). Subgrouping Patients with Nonspecific Low Back Pain: Hope or Hype? *The journal of orthopaedic and sports physical therapy, 47*(2), 44-48. doi:10.2519/jospt.2017.0602

Scaia, V., Baxter, D., & Cook, C. (2012). The pain provocation-based straight leg raise test for diagnosis of lumbar disc herniation, lumbar radiculopathy, and/or sciatica: A systematic review of clinical utility. *Journal of Back and Musculoskeletal Rehabilitation*, *25*, 215–223. doi:10.3233/BMR-2012-0339

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*, 671–675. doi:10.1038/nmeth.2089

Schwarzer, A. C., Derby, R., Aprill, C. N., Fortin, J., Kine, G., & Bogduk, N. (1994). Pain from the lumbar zygapophysial joints: A test of two models. *Journal of Spinal Disorders*, *7*, 331–336.

Sehgal, N., Dunbar, E. E., Shah, R. V., & Colson, J. (2007). Systematic review of diagnostic utility of facet (zygapophysial) joint injections in chronic spinal pain: An update. *Pain Physician*, *10*(1), 213.

Seising, R. (2006). From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis. *Artificial Intelligence in Medicine, 38*(3), 237–256. doi:10.1016/j.artmed.2006.06.004

Shiri, R., Karppinen, J., Leino-Arjas, P., Solovieva, S., & Viikari-Juntura, E. (2010). The association between obesity and low back pain: A meta-analysis. *American Journal of Epidemiology*, *171*, 135–154. doi:10.1093/aje/kwp356

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. doi:10.1037//0033-2909.86.2.420

Simopoulos, T. T., Manchikanti, L., Singh, V., Gupta, S., Hameed, H., Diwan, S., & Cohen, S. P. (2012). A systematic evaluation of prevalence and diagnostic accuracy of sacroiliac joint interventions. *Pain Physician*, *15*(3), E305–344.

Sivayogam, A., & Banerjee, A. (2011). Diagnostic performance of clinical tests for sacroiliac joint pain. *Physical Therapy Reviews*, *16*, 462–467. doi:10.1179/1743288X11Y.0000000036

Stanos, S. (2012). Focused review of interdisciplinary pain rehabilitation programs for chronic pain management. *Current Pain and Headache Reports*, *16*(2), 147–152. doi:10.1007/s11916-012-0252-4

Stojanovic, M. P., & Engel, A. J. (2015). An expedited diagnostic algorithm for lumbosacral pain: 'The best likelihood scenario approach'. *Pain Medicine*, *16*, 2271–2276. doi:10.1111/pme.12839

Stout, A. (2010). Discography. *Physical Medicine and Rehabilitation Clinics of North America*, *21*(4), 859–867. doi:10.1016/j.pmr.2010.07.002

Streicher, H., Mätzold, F., Hamilton, C., & Wagner, P. (2014). Comparison of group motor control training versus individual training for people suffering from back pain. *Journal of Bodywork and Movement Therapies*, *18*, 489–496. doi:10.1016/j.jbmt.2013.12.006

Tahan, N., Rasouli, O., Arab, A. M., Khademi, K., & Samani, E. N. (2014). Reliability of the ultrasound measurements of abdominal muscles activity when activated with and without pelvic floor muscles contraction. *Journal of Back & Musculoskeletal Rehabilitation*, *27*(3), 339–347. doi:10.3233/BMR-130453

Tawa, N., Rhoda, A., & Diener, I. (2016). Accuracy of magnetic resonance imaging in detecting lumbo-sacral nerve root compromise: A systematic literature review. *BMC Musculoskeletal Disorders*, *17*. doi:10.1186/s12891-016-1236-z

Tesh, K. M., Shaw Dunn, J., & Evans, J. H. (1987). The abdominal muscles and vertebral stability. *Spine*, *12*, 501–508.

Teyhen, D. S., Miltenberger, C. E., Deiters, H. M., Toro, Y. M., Pulliam, J. N., Childs, J. D., … Flynn, T. W. (2005). The use of ultrasound imaging of the abdominal drawing-in maneuver in subjects with low back pain. *Journal of Orthopaedic & Sports Physical Therapy*, *35*, 346–355. doi:10.2519/jospt.2005.35.6.346

Thanh Le, N. T., Robinson, J., & Lewis, S. J. (2015). A study of student radiographers' learning experiences in imaging obese patients. *Journal of Medical Imaging and Radiation Sciences*, *46*(3), S61–S68.e61. doi:10.1016/j.jmir.2015.05.002

Tousignant, M., Poulin, L., Marchand, S., Viau, A., & Place, C. (2005). The
    modified–modified Schöber test for range of motion assessment of lumbar
    flexion in patients with low back pain: A study of criterion validity, intra- and
    inter-rater reliability and minimum metrically detectable change. *Disability
    and Rehabilitation*, *27*, 553–559. doi:10.1080/09638280400018411

Tracey, I., & Bushnell, C. M. (2009). How neuroimaging studies have challenged us
    to rethink: Is chronic pain a disease? *The Journal of Pain*, *10*(11), 1113–
    1120. doi:http://dx.doi.org/10.1016/j.jpain.2009.09.001

Trudelle-Jackson, E., Sarvaiya-Shah, S. A., & Wang, S. S. (2008). Interrater
    reliability of a movement impairment-based classification system for lumbar
    spine syndromes in patients with chronic low back pain. *Journal of
    Orthopaedic & Sports Physical Therapy*, *38*, 371–376.
    doi:10.2519/jospt.2008.2760

Tsao, H., Danneels, L. A., & Hodges, P. W. (2011). ISSLS prize winner: Smudging
    the motor brain in young adults with recurrent low back pain. *Spine, 36*,
    1721–1727. doi:10.1097/BRS.0b013e31821c4267

Tucker, N., Reid, D., & McNair, P. (2007). Reliability and measurement error of
    active knee extension range of motion in a modified slump test position: A
    pilot study. *Journal of Manual & Manipulative Therapy*, *15*, E85–E91.

Unsgaard-Tøndel, M., Lund Nilsen, T. I., Magnussen, J., & Vasseljen, O. (2012). Is
    activation of transversus abdominis and obliquus internus abdominis
    associated with long-term changes in chronic low back pain? A prospective
    study with 1-year follow-up. *British Journal of Sports Medicine*, *46*, 729–
    734. doi:10.1136/bjsm.2011.085506

van der Hulst, M., Vollenbroek-Hutten, M. M. R., & Ijzerman, M. J. (2005). A

    systematic review of sociodemographic, physical, and psychological

    predictors of multidisciplinary rehabilitation—or, back school treatment

    outcome in patients with chronic low back pain. *Spine*, *30*, 813–825.

van der Wurff, P., Hagmeijer, R. H. M., & Meyne, W. (2000). Clinical tests of the

    sacroiliac joint: A systematic methodological review. Part 1: Reliability.

    Manual Therapy, *5*, 30–36. doi:10.1054/math.1999.0228

Vibe Fersum, K., O'Sullivan, P. B., Kvåle, A., & Skouen, J. S. (2009). Inter-

    examiner reliability of a classification system for patients with non-specific

    low back pain. *Manual Therapy*, *14*, 555–561.

    doi:10.1016/j.math.2008.08.003

Viniol, A., Jegan, N., Leonhardt, C., Brugger, M., Strauch, K., Barth, J., … Becker,

    A. (2013). Differences between patients with chronic widespread pain and

    local chronic low back pain in primary care—a comparative cross-sectional

    analysis. *BMC Musculoskeletal Disorders*, *14*, 1–9. doi:10.1186/1471-2474-

    14-351

von Baeyer, C. L., Lin, V., Seidman, L. C., Tsao, J. C. I., & Zeltzer, L. K. (2011).

    Pain charts (body maps or manikins) in assessment of the location of

    pediatric pain. *Pain Management*, *1*, 61–68. doi:10.2217/pmt.10.2

Wade, R., Corbett, M., & Eastwood, A. (2013). Quality assessment of comparative

    diagnostic accuracy studies: Our experience using a modified version of the

    QUADAS-2 tool. *Research Synthesis Methods*, *4*, 280–286.

    doi:10.1002/jrsm.1080

Wand, B. M., Parkitny, L., O'Connell, N. E., Luomajoki, H., McAuley, J. H.,

    Thacker, M., & Moseley, G. L. (2011). Cortical changes in chronic low back

pain: Current state of the art and implications for clinical practice. *Manual*

Therapy, 16, 15–20. doi:10.1016/j.math.2010.06.008

Wassenaar, M., van Rijn, R. M., van Tulder, M. W., Verhagen, A. P., van der Windt,

D. A. W. M., Koes, B. W., … Ostelo, R. W. J. G. (2012). Magnetic resonance

imaging for diagnosing lumbar spinal pathology in adult patients with low

back pain or sciatica: A diagnostic systematic review. *European Spine*

Journal, *21*, 220–227. doi:10.1007/s00586-011-2019-8

Watanabe, S., Kobara, K., Yoshimura, Y., Osaka, H., & Ishida, H. (2014). Influence

of trunk muscle co-contraction on spinal curvature during sitting. *Journal of*

Back and Musculoskeletal Rehabilitation, *27*, 55–61. doi:10.3233/BMR-

130419

Werneke, M. W., Hart, D. L., Cutrone, G., Oliver, D., McGill, T., Weinberg, J., …

Ward, J. (2011). Association between directional preference and

centralization in patients with low back pain. *Journal of Orthopaedic &*

Sports Physical Therapy, *41*, 22–31. doi:10.2519/jospt.2011.3415

Whiting, P. F., Rutjes, A. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reltsma, J.

B., … Bossuyt, P. M. M. (2011). Quadas-2: A revised tool for the quality

assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*,

529–536. doi:10.7326/0003-4819-155-8-201110180-00009

Whittaker, J. L., & Stokes, M. (2011). Ultrasound imaging and muscle function.

Journal of Orthopaedic & Sports Physical Therapy, *41*, 572–580.

doi:10.2519/jospt.2011.3682

Whittaker, J. L., Warner, M. B., & Stokes, M. J. (2010). Ultrasound imaging

transducer motion during clinical maneuvers: Respiration, active straight leg

raise test and abdominal drawing in. *Ultrasound in Medicine & Biology*, *36*,

   1288–1297. doi:10.1016/j.ultrasmedbio.2010.04.020

Whittle, C. J., Flavell, C. A., & Gordon, S. J. (2017). Methodological consistency

   and measurement reliability of transversus abdominis real time ultrasound

   imaging in chronic low back pain populations: A systematic review. *Physical*

   Therapy Reviews, *22*, 48–59. doi:10.1080/10833196.2017.1287151

Widerström, B., Olofsson, N., Arvidsson, I., Harms-Ringdahl, K., & Larsson, U. E.

   (2012). Inter-examiner reliability of a proposed decision-making treatment

   based classification system for low back pain patients. *Manual Therapy*, *17*,

   164–171. doi:10.1016/j.math.2011.12.009

Wilde, V. E., Ford, J. J., & McMeeken, J. J. (2007). Indicators of lumbar

   zygapophyseal joint pain: Survey of an expert panel with the Delphi

   technique. *Physical Therapy*, *87*, 1348–1361.

Wilson, L., Hall, H., McIntosh, G., & Melles, T. (1999). Intertester reliability of a

   low back pain classification system. *Spine*, *24*, 248–254.

   doi:10.1097/00007632-199902010-00009

Womersley, L., & May, S. (2006). Sitting posture of subjects with postural backache.

   Journal of Manipulative and Physiological Therapeutics, *29*, 213–218.

   doi:10.1016/j.jmpt.2006.01.002

Wong, A. Y. L., Parent, E. C. Funabashi, M., Stanton, T. R., & Kawchuk, G. N.

   (2013). Do various baseline characteristics of transversus abdominis and

   lumbar multifidus predict clinical outcomes in non-specific low back pain? A

   systematic review. *Pain, 154*, 2589-2602.

Wu, G., & Cavanagh, P. R. (1995). ISB recommendations for standardization in the reporting of kinematic data. *Journal of Biomechanics*, *28*, 1257–1261. doi:10.1016/0021-9290(95)00017-C

Wu, T., Zhao, W., Dong, Y., Song, H., & Li, J. (2016). Effectiveness of ultrasound-guided versus fluoroscopy or computed tomography scanning guidance in lumbar facet joint injections in adults with facet joint syndrome: A meta-analysis of controlled trials. *Archives of Physical Medicine and Rehabilitation*, *97*, 1558–1563. doi:10.1016/j.apmr.2015.11.013

Appendices

Appendix 1: Self-Reported Questionnaires (with permission to reproduce)

1A: Oswestry Disability Index

This administrative form
has been removed

This administrative form
has been removed

232

This administrative form
has been removed

234

235

236

This administrative form
has been removed

238

This administrative form
has been removed

240

Oswestry Disability Index (ODI)

MAPI RESEARCH
TRUST
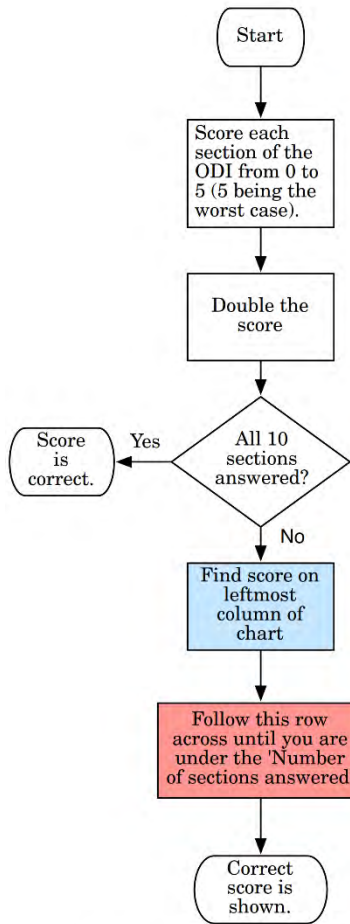Education
Information
Dissemination

SCALING AND SCORING OF THE

Oswestry Disability Index (ODI)

Mapi Research Trust
27 rue de la Villette
69003 Lyon
France
Phone: +33 (0) 4 72 13 65 75
Fax: +33 (0) 4 72 13 66 82

Contact:
Marie-Pierre Emery
E-mail: mpemery@mapigroup.com

Jeremy Fairbank
Orthopaedic Surgery
Nuffield Orthopaedic Centre
Oxford OX3 7LD
UK

241

ODI made easy

	Number of sections answered								
	9	8	7	6	5	4	3	2	1
0	0	0	0	0	0	0	0	0	0
2	2	2	3	3	4	5	7	10	20
4	4	5	6	7	8	10	13	20	40
6	7	8	9	10	12	15	20	30	60
8	9	10	11	13	16	20	27	40	80
10	11	12	14	17	20	25	33	50	100
12	13	15	17	20	24	30	40	60	
14	16	18	20	23	28	35	47	70	
16	18	20	23	27	32	40	53	80	
18	20	22	26	30	36	45	60	90	
20	22	25	29	33	40	50	67	100	
22	24	28	31	37	44	55	73		
24	27	30	34	40	48	60	80		
26	29	32	37	43	52	65	87		
28	31	35	40	47	56	70	93		
30	33	38	43	50	60	75	100		
32	36	40	46	53	64	80			
34	38	42	49	57	68	85			
36	40	45	51	60	72	90			
38	42	48	54	63	76	95			
40	44	50	57	67	80	100			
42	47	52	60	70	84				
44	49	55	63	73	88				
46	51	58	66	77	92				
48	53	60	69	80	96				
50	56	62	71	83	100				
52	58	65	74	87					
54	60	68	77	90					
56	62	70	80	93					
58	64	72	83	97					
60	67	75	86	100					
62	69	78	89						
64	71	80	91						
66	73	82	94						
68	76	85	97						
70	78	88	100						
72	80	90							
74	82	92							
76	84	95							
78	87	98							
80	89	100							
82	91								
84	93								
86	96								
88	98								
90	100								

Flowchart:

Start → Score each section of the ODI from 0 to 5 (5 being the worst case). → Double the score → All 10 sections answered?
- Yes → Score is correct.
- No → Find score on leftmost column of chart → Follow this row across until you are under the 'Number of sections answered' → Correct score is shown.

Example
Score from questionnaire = 20
Doubled score = 40
Sections answered = 9
Correct score = 44%

ODI version 2.1a

This questionnaire is designed to give us information as to how your back (or leg) trouble affects your ability to manage in everyday life.
Please answer every section. Mark one box only in each section that most closely describes you today.

Section 1 - Pain intensity

- ☐ I have no pain at the moment.
- ☐ The pain is very mild at the moment.
- ☐ The pain is moderate at the moment.
- ☐ The pain is fairly severe at the moment.
- ☐ The pain is very severe at the moment.
- ☐ The pain is the worst imaginable at the moment.

Section 2 - Personal care (washing, dressing, etc.)

- ☐ I can look after myself normally without causing extra pain.
- ☐ I can look after myself normally but it is very painful.
- ☐ It is painful to look after myself and I am slow and careful.
- ☐ I need some help but manage most of my personal care.
- ☐ I need help every day in most aspects of self care.
- ☐ I do not get dressed, wash with difficulty and stay in bed.

Section 3 - Lifting

- ☐ I can lift heavy weights without extra pain.
- ☐ I can lift heavy weights but it gives extra pain.
- ☐ Pain prevents me from lifting heavy weights off the floor but I can manage if they are conveniently positioned, e.g. on a table.
- ☐ Pain prevents me from lifting heavy weights but I can manage light to medium weights if they are conveniently positioned.
- ☐ I can lift only very light weights.
- ☐ I cannot lift or carry anything at all.

Section 4 - Walking

- ☐ Pain does not prevent me walking any distance.
- ☐ Pain prevents me walking more than one mile.
- ☐ Pain prevents me walking more than a quarter of a mile.
- ☐ Pain prevents me walking more than 100 yards.

- ☐ I can only walk using a stick or crutches.
- ☐ I am in bed most of the time and have to crawl to the toilet.

Section 5 - Sitting

- ☐ I can sit in any chair as long as I like.
- ☐ I can sit in my favourite chair as long as I like.
- ☐ Pain prevents me from sitting for more than 1 hour.
- ☐ Pain prevents me from sitting for more than half an hour.
- ☐ Pain prevents me from sitting for more than 10 minutes.
- ☐ Pain prevents me from sitting at all.

Section 6 - Standing

- ☐ I can stand as long as I want without extra pain.
- ☐ I can stand as long as I want but it gives me extra pain.
- ☐ Pain prevents me from standing for more than 1 hour.
- ☐ Pain prevents me from standing for more than half an hour.
- ☐ Pain prevents me from standing for more than 10 minutes.
- ☐ Pain prevents me from standing at all.

Section 7 - Sleeping

- ☐ My sleep is never disturbed by pain.
- ☐ My sleep is occasionally disturbed by pain.
- ☐ Because of pain I have less than 6 hours sleep.
- ☐ Because of pain I have less than 4 hours sleep.
- ☐ Because of pain I have less than 2 hours sleep.
- ☐ Pain prevents me from sleeping at all.

Section 8 - Sex life (if applicable)

- ☐ My sex life is normal and causes no extra pain.
- ☐ My sex life is normal but causes some extra pain.
- ☐ My sex life is nearly normal but is very painful.
- ☐ My sex life is severely restricted by pain.
- ☐ My sex life is nearly absent because of pain.
- ☐ Pain prevents any sex life at all.

Section 9 - Social life

- ☐ My social life is normal and causes me no extra pain.

- ☐ My social life is normal but increases the degree of pain.

- ☐ Pain has no significant effect on my social life apart from limiting my more energetic interests, e.g. sport, etc.

- ☐ Pain has restricted my social life and I do not go out as often.

- ☐ Pain has restricted social life to my home.

- ☐ I have no social life because of pain.

Section 10 - Travelling

- ☐ I can travel anywhere without pain.

- ☐ I can travel anywhere but it gives extra pain.

- ☐ Pain is bad but I manage journeys over two hours.

- ☐ Pain restricts me to journeys of less than one hour.

- ☐ Pain restricts me to short necessary journeys under 30 minutes.

- ☐ Pain prevents me from travelling except to receive treatment

---

Result

Your ODI = ☐ %

1B: Roland Morris Disability Questionnaire

Roland, M. O., & Fairbank, J. C. T. (2000). The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine, 25*(24), 3115–3124.

This administrative form
has been removed

Roland-Morris Low Back Pain
And Disability Questionnaire (RMQ)

Source: Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther*. 1996 Apr;76(4):359-65; discussion 66-8.

The Roland-Morris Questionnaire (RMQ) is a self-administered disability measure in which greater levels of disability are reflected by higher numbers on a 24-point scale. The RMQ has been shown to yield reliable measurements, which are valid for inferring the level of disability, and to be sensitive to change over time for groups of patients with low back pain.

Scoring instructions

The patient is instructed to put a mark next to each appropriate statement. Add up the total number of marked statements to get a patient's score.

Interpretation of scores

Roland and Morris did not provide descriptions of the varying degrees of disability (eg, 40%-60% is severe disability). Clinical improvement over time can be graded based on the analysis of serial questionnaire scores. If, for example, at the beginning of treatment, a patient's score was 12 and, at the conclusion of treatment, their score was 2 (10 points of improvement), we would calculate an 83% (10/12 x 100) improvement.

Roland-Morris Low Back Pain and Disability Questionnaire (RMQ)

Instructions

Patient name:_____ File #:_____ Date:_____

Please read instructions: When your back hurts, you may find it difficult to do some of the things you normally do. Mark only the sentences that describe you today.

- ☐ I stay at home most of the time because of my back.
- ☐ I change position frequently to try to get my back comfortable.
- ☐ I walk more slowly than usual because of my back.
- ☐ Because of my back, I am not doing any jobs that I usually do around the house.
- ☐ Because of my back, I use a handrail to get upstairs.
- ☐ Because of my back, I lie down to rest more often.
- ☐ Because of my back, I have to hold on to something to get out of an easy chair.
- ☐ Because of my back, I try to get other people to do things for me.
- ☐ I get dressed more slowly than usual because of my back.
- ☐ I only stand up for short periods of time because of my back.
- ☐ Because of my back, I try not to bend or kneel down.
- ☐ I find it difficult to get out of a chair because of my back.
- ☐ My back is painful almost all of the time.
- ☐ I find it difficult to turn over in bed because of my back.
- ☐ My appetite is not very good because of my back.
- ☐ I have trouble putting on my socks (or stockings) because of the pain in my back.
- ☐ I can only walk short distances because of my back pain.
- ☐ I sleep less well because of my back.
- ☐ Because of my back pain, I get dressed with the help of someone else.
- ☐ I sit down for most of the day because of my back.
- ☐ I avoid heavy jobs around the house because of my back.
- ☐ Because of back pain, I am more irritable and bad tempered with people than usual.
- ☐ Because of my back, I go upstairs more slowly than usual.
- ☐ I stay in bed most of the time because of my back.

1C: The Modified Somatic Pain Questionnaire

Main, C. J. (1983). The Modified Somatic Pain Questionnaire (MSPQ). *Journal of Psychosomatic Research*, *27*(6), 503-514. DOI 10.1016/0022-3999(83)90040-5

This administrative form
has been removed

Please describe how you have felt during the PAST WEEK by marking a check mark (☐) in the appropriate box. Please answer all questions. Do not think too long before answering.

	Not at all	A little/ slightly	A great deal/ quite a bit	Extremely/ could not have been worse
Feeling hot all over				
Sweating all over				
Dizziness				
Blurring of vision				
Feeling faint				
Nausea				
Pain or ache in stomach				
Stomach churning				
Mouth becoming dry				
Muscles in neck aching				
Legs feeling weak				
Muscles twitching or jumping				
Tense feeling across forehead				

Response for each somatic perception	
Not at all	0
A little/slightly	1
A great deal/quite a bit	2
Extremely/could not have been worse	3

Interpretation:
-Minimum score 0   - Maximum score 39
-Higher the score the more marked the somatic symptoms

-At risk < 12   - Distressed-somatic ≥ 12

1D: The 11 Point Visual Analogue Scale

Huskisson, E. C. (1974). Measurement of Pain. *The Lancet*, *304*(7889), 1127-1131.

Visual Analog Scale (VAS)†

No
Pain

Pain As Bad
As It Could
Possibly Be
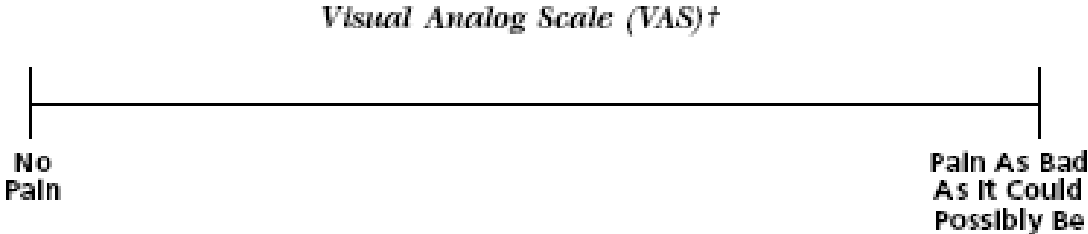
Appendix 2: QUADAS Background Document (with permission to reproduce)

(http://www.bristol.ac.uk/medialibrary/sites/quadas/migrated/documents/background
-doc.pdf)

This administrative form
has been removed

254

QUADAS-2: Background Document

QUADAS-2

QUADAS-2 is designed to assess the quality of primary diagnostic accuracy studies; it is not designed to replace the data extraction process of the review and should be applied in addition to extracting primary data (e.g. study design, results etc) for use in the review. It consists of four key domains covering patient selection, index test, reference standard, and flow of patients through the study and timing of the index test(s) and reference standard ("flow and timing") (Table 1). The tool is completed in four phases: 1) state the review question; 2) develop review specific guidance; 3) review the published flow diagram for the primary study or construct a flow diagram if none is reported; 4) judgement of bias and applicability. Each domain is assessed in terms of the *risk of bias* and the first three are also assessed in terms of *concerns regarding applicability*. To help reach a judgement on the risk of bias, *signalling questions* are included. These flag aspects of study design related to the potential for bias and aim to help reviewers make risk of bias judgements.

Phase 1: Review Question

Review authors are first asked to report their systematic review question in terms of patients, index test(s), and reference standard and target condition. As the accuracy of a test may depend on where in the diagnostic pathway it will be used, review authors are asked to describe patients in terms of setting, intended use of the index test, patient presentation and prior testing.(1;2)

Phase 2: Review Specific Tailoring (Figure 1)

It is essential to tailor QUADAS-2 to each review by adding or omitting signalling questions and developing review-specific guidance on how to assess each signalling question and use this information to judge the risk of bias. The first step is to consider whether any signalling question does not apply to the review or whether any specific issues for the review are not adequately covered by the core signalling questions. For example, for a review of an objective index test it may be appropriate to omit the signalling question relating to blinding of the test interpreter to results of the reference standard. Review authors should avoid

complicating the tool by adding too many signalling questions. Once tool content has been agreed, review-specific rating guidance should be developed. The tool should be piloted independently by at least two people. If agreement is good, the tool can be used to rate all included studies. If agreement is poor, further refinement may be needed.

Figure 1: Process for tailoring QUADAS-2 to your systematic review
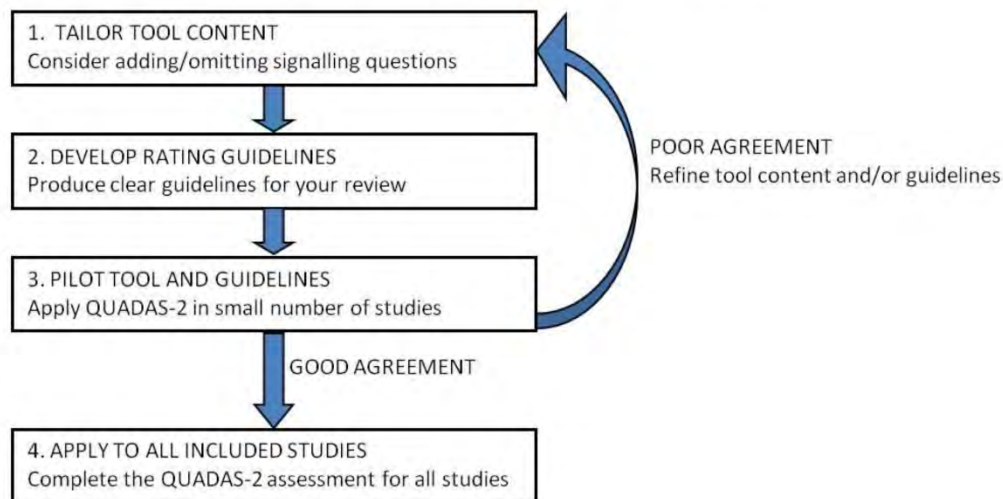
Phase 3: Flow Diagram

The next stage is to review the published flow diagram for the primary study or to draw one if none is reported or the published diagram is not adequate. The flow diagram will facilitate judgments of risk of bias, and should provide information about the method of recruitment of patients (e.g. based on a consecutive series of patients with specific symptoms suspected of having the target condition, or of cases and controls), the order of test execution, and the number of patients undergoing the index test and the reference standard. A hand drawn diagram is sufficient as this step does not need to be reported as part of the QUADAS-2 assessment. Figure 2 shows an example based on a primary study of B type natriuretic peptide for the diagnosis of heart failure.

Figure 2: Flowchart based on diagnostic cohort study of BNP for diagnosing heart failure
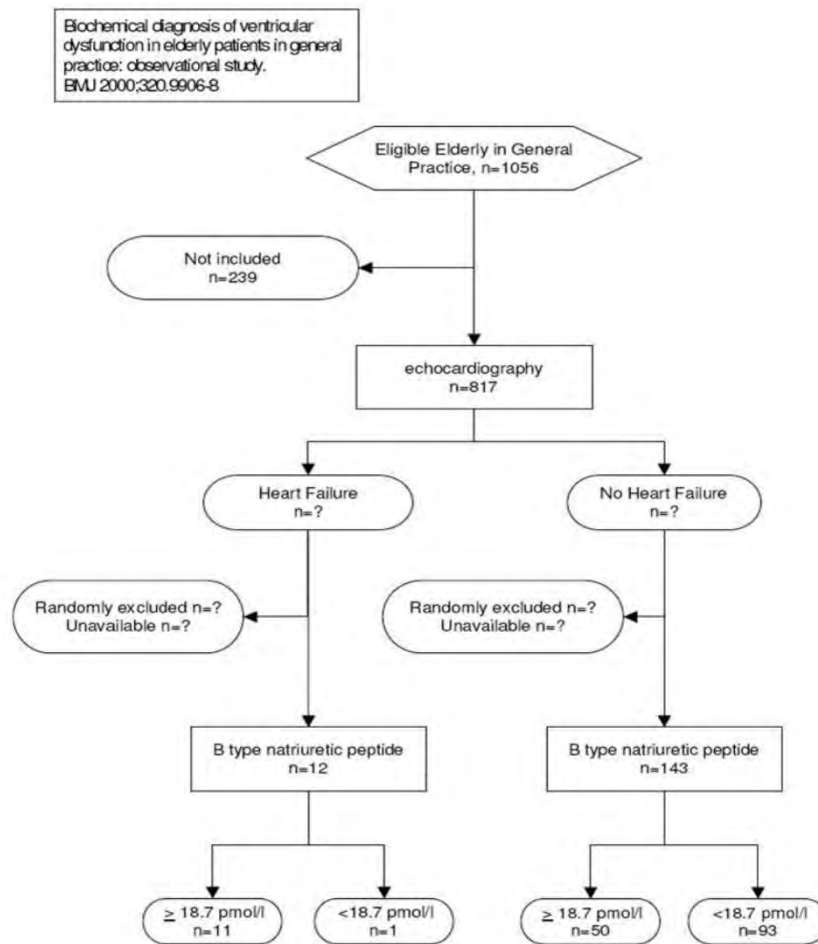
Phase 4: Judgments on bias and applicability

Risk of bias

The first part of each domain concerns bias and comprises three sections: 1) information used to support the risk of bias judgment, 2) signalling questions, and 3) judgment of risk of bias. By recording the information used to reach the judgment ("support for judgment"), we aim to make the rating transparent and facilitate discussion between review authors completing assessments independently.(3) The additional signalling questions are included

to assist judgments. They are answered as "yes", "no", or "unclear", and are phrased such that "yes" indicates low risk of bias.

Risk of bias is judged as "low", "high", or "unclear". If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias. Review authors then need to use the guidelines developed in phase 2 to judge risk of bias. The "unclear" category should be used only when insufficient data are reported to permit a judgment.

Applicability

Applicability sections are structured in a similar way to the bias sections, but do not include signalling questions. Review authors are asked to record the information on which the judgment of applicability is made and then to rate their concern that the study does not match the review question. Concerns regarding applicability are rated as "low", "high" or "unclear". Applicability judgments should refer to the first phase, where the review question was recorded. Again, the "unclear" category should only be used when insufficient data are reported.

The following sections provide brief explanations of the signalling questions and risk of bias/concerns regarding applicability questions for each domain.

DOMAIN 1: PATIENT SELECTION

Risk of bias: Could the selection of patients have introduced bias?

Signalling question 1: Was a consecutive or random sample of patients enrolled?

Signalling question 2: Was a case-control design avoided?

Signalling question 3: Did the study avoid inappropriate exclusions?

A study should ideally enrol all consecutive, or a random sample of, eligible patients with suspected disease – otherwise there is potential for bias. Studies that make inappropriate exclusions, e.g. excluding "difficult to diagnose" patients, may result in overoptimistic estimates of diagnostic accuracy. In a review of anti-CCP antibodies for the diagnosis of rheumatoid arthritis, we found that some studies enrolled consecutive patients who had confirmed diagnoses. These studies showed greater sensitivity of the anti-CCP test than

studies that included patients with suspected disease but in whom the diagnosis had not been confirmed – "difficult to diagnose" patients.(4) Similarly, studies enrolling patients with known disease and a control group without the condition may exaggerate diagnostic accuracy.(5;6) Exclusion of patients with "red flags" for the target condition, who may be easier to diagnose, may lead to underestimation of diagnostic accuracy.

Applicability: Are there concerns that the included patients and setting do not match the review question?

There may be concerns regarding applicability if patients included in the study differ, compared to those targeted by the review question, in terms of severity of the target condition, demographic features, presence of differential diagnosis or co-morbidity, setting of the study and previous testing protocols. For example, larger tumours are more easily seen with imaging tests than smaller ones, and larger myocardial infarctions lead to higher levels of cardiac enzymes than small infarctions making them easier to detect and so increasing estimates of sensitivity.(7)

DOMAIN 2: INDEX TEST

Risk of Bias: Could the conduct or interpretation of the index test have introduced bias?
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?

This item is similar to "blinding" in intervention studies. Interpretation of index test results may be influenced by knowledge of the reference standard.(6) The potential for bias is related to the subjectivity of index test interpretation and the order of testing. If the index test is always conducted and interpreted prior to the reference standard,this item can be rated "yes".

Signalling question 2: If a threshold was used, was it pre-specified?

Selecting the test threshold to optimise sensitivity and/or specificity may lead to overoptimistic estimates of test performance, which is likely to be poorer in an independent sample of patients in whom the same threshold is used.(8)

Applicability: Are there concerns that the index test, its conduct, or interpretation differ from the review question?

Variations in test technology, execution, or interpretation may affect estimates of its diagnostic accuracy. If index tests methods vary from those specified in the review question there may be concerns regarding applicability. For example, a higher ultrasound transducer frequency has been shown to improve sensitivity for the evaluation of patients with abdominal trauma.(9)

DOMAIN 3: REFERENCE STANDARD

Risk of Bias: Could the reference standard, its conduct, or its interpretation have introduced bias?

Signalling question 1: Is the reference standard likely to correctly classify the target condition?

Estimates of test accuracy are based on the assumption that the reference standard is 100% sensitive and specific disagreements between the reference standard and index test are assumed to result from incorrect classification by the index test.(10;11)

Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?

This item is similar to the signalling question related to interpretation of the index test. Potential for bias is related to the potential influence of prior knowledge on the interpretation of the reference standard.(6)

Applicability: Are there concerns that the target condition as defined by the reference standard does not match the question?

The reference standard may be free of bias but the target condition that it defines may differ from the target condition specified in the review question. For example, when defining urinary tract infection the reference standard is generally based on specimen culture but the threshold above which a result is considered positive may vary.(12)

DOMAIN 4: FLOW AND TIMING

Risk of Bias: Could the patient flow have introduced bias?

Signalling question 1: Was there an appropriate interval between index test and reference standard?

260

Ideally results of the index test and reference standard are collected on the same patients at the same time. If there is a delay or if treatment is started between index test and reference standard, misclassification may occur due to recovery or deterioration of the condition. The length of interval leading to a high risk of bias will vary between conditions. A delay of a few days may not be a problem for chronic conditions, while for acute infectious diseases a short delay may be important. Conversely, when the reference standard involves follow-up a minimum follow-up period may be required to assess the presence or absence of the target condition. For example, for the evaluation of magnetic resonance imaging for the early diagnosis of multiple sclerosis, a minimum follow-up period of around 10 years is required to be confident that all patients who will go on to fulfil diagnostic criteria for multiple sclerosis will have done so.(13)

Signalling question 2: Did all patients receive the same reference standard?
Verification bias occurs when not all of the study group receive confirmation of the diagnosis by the same reference standard. If the results of the index test influence the decision on whether to perform the reference standard or which reference standard is used, estimated diagnostic accuracy may be biased.(5;14) For example, a study evaluating the accuracy of the D-dimer test for the diagnosis of pulmonary embolism carried out ventilation perfusion scans (reference standard 1) in those testing positive and used clinical follow-up to determine whether or not those testing negative had a pulmonary embolism (reference standard 2). This may result in misclassifying some of the false negatives as true negatives as some patients who had a pulmonary embolism but were index test negative may be missed by clinical follow-up and so be classified as not having a pulmonary embolism. This misclassification will overestimate sensitivity and specificity.

Signalling question 3: Were all patients included in the analysis?
All patients who were recruited into the study should be included in the analysis.(15) There is a potential for bias if the number of patients enrolled differs from the number of patients included in the 2x2 table of results, for example because patients lost to follow-up differ systematically from those who remain.

261

Incorporating QUADAS-2 assessments in diagnostic accuracy reviews

We emphasise that QUADAS-2 should not be used to generate a summary "quality score", because of the well-known problems associated with such scores.(16;17) If a study is judged as "low" on all domains relating to bias or applicability then it is appropriate to have an overall judgment of "low risk of bias" or "low concern regarding applicability" for that study. If a study is judged "high" or "unclear" on one or more domains then it may be judged "at risk of bias" or as having "concerns regarding applicability".

At minimum, reviews should present a summary of the results of the QUADAS-2 assessment for all included studies. This could include summarising the number of studies that found low, high or unclear risk of bias/concerns regarding applicability for each domain. If studies are found to consistently rate well or poorly on particular signalling questions then reviewers may choose to highlight these. Tabular (Table) and graphical (Figure 3) displays are helpful to summarise QUADAS-2 assessments.

Table: Suggested tabular presentation for QUADAS-2 results

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Study 1	☺	☺	☺	☺	☹	☺	☺
Study 2	☺	☺	☺	☺	☹	☺	☺
Study 3	☹	☹	☺	☺	☹	☺	☺
Study 4	☹	☹	☺	☺	☹	☺	☺
Study 5	☹	?	☺	☺	☹	☺	☺
Study 6	☹	?	☺	☺	☹	?	☺
Study 7	☹	?	☺	☺	☹	☺	☺
Study 8	☹	?	☺	☺	☹	?	☺
Study 9	☹	?	☺	☺	☹	☺	☺
Study 10	☹	?	☺	☹	☹	☺	☺
Study 11	☺	?	☺	☹	☺	☺	☺

☺ Low Risk   ☹ High Risk   ? Unclear Risk

Figure 3: Suggested Graphical Display for QUADAS-2 results
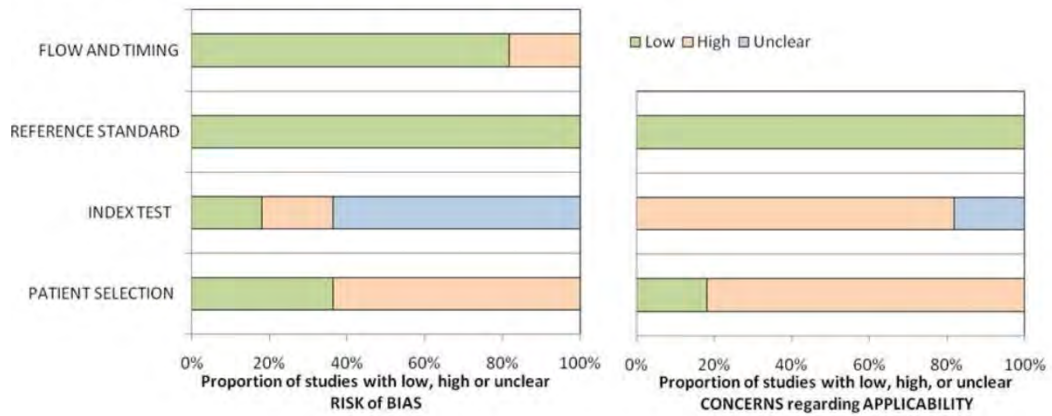
Review authors may choose to restrict the primary analysis so that only studies at low risk of bias and/or low concern regarding applicability for all or specified domains are included. It may be appropriate to restrict inclusion to the review based on similar criteria, but it is often preferable to review all relevant evidence and then investigate possible reasons for heterogeneity.(13;18) Subgroup and or sensitivity analysis can be conducted by investigating how estimates of accuracy of the index test vary between studies rated as high, low, or unclear on all or selected domains. Domains or signalling questions can be included as items in meta-regression analyses, to investigate their association with estimated accuracy.

Website

The QUADAS website (www.quadas.org) contains QUADAS-2, information on training, a bank of additional signalling questions, more detailed guidance for each domain, examples of completed QUADAS-2 assessments, and downloadable resources including a Microsoft Access™ database for data extraction, an Excel™ spreadsheet to produce graphical displays of results, and templates for Word™ tables to summarise results.

References

(1) Bossuyt PM, Leeflang MMG. Chapter 6: Developing Criteria for Including Studies. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration; 2009.

(2) Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008; 149(12):889-897.

(3) Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. BMJ. In press 2011.

(4) Whiting PF, Smidt N, Sterne JA, Harbord R, Burton A, Burke M et al. Systematic review: accuracy of anti-citrullinated Peptide antibodies for diagnosing rheumatoid arthritis. Ann Intern Med 2010; 152(7):456-464.

(5) Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282(11):1061-1066.

(6) Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004; 140(3):189-202.

(7) Reitsma J, Rutjes A, WP, Vlassov V, Leeflang M, Deeks J. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration; 2009.

(8) Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clinical Chemistry 2008; 54(4):729-737.

(9) Stengel D, Bauwens K, Rademacher G, Mutze S, Ekkernkamp A. Association between compliance with methodological standards of diagnostic research and reported test accuracy: meta-analysis of focused assessment of US for trauma. Radiology 2005; 236(1):102-111.

(10) Biesheuvel C, Irwig L, Bossuyt P. Observed differences in diagnostic test accuracy between patient subgroups: is it real or due to reference standard misclassification? Clin Chem 2007; 53(10):1725-1729.

(11) van Rijkom HM, Verdonschot EH. Factors involved in validity measurements of diagnostic tests for approximal caries--a meta-analysis. Caries Research 1995; 29(5)):364-70.

(12) Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J et al. Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model. Health Technol Assess 2006; 10(36):iii-xiii, 1.

(13) Whiting P, Harbord R, Main C, Deeks JJ, Filippini G, Egger M et al. Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis: systematic review. BMJ 2006; 332(7546):875-884.

(14) Rutjes A, Reitsma J, Di NM, Smidt N, Zwinderman A, Van RJ et al. Bias in estimates of diagnostic accuracy due to shortcomings in design and conduct: empirical evidence [abstract]. XI Cochrane Colloquium: Evidence, Health Care and Culture; 2003 Oct 26 31; Barcelona, Spain 2003;45.

(15) Macaskill P, Gatsonis C, Deeks JJ, Harbord R, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration; 20010.

(16) Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282(11):1054-1060.

(17) Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol 2005; 5:19.

(18) Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Medical Research Methodology 2006; 6:9.

Appendix 3: Published Papers (with permission to reproduce)

Flavell, C. A., Gordon, S., Marshman, L & Watt, K. (2014). Inter-rater reliability of

    classification systems in chronic low back pain populations. *Physical*

    Therapy reviews, *19*(3), 204-212. DOI 10.1179/1743288X13Y.0000000131

This administrative form
has been removed

Content has been removed
due to copyright restrictions

270

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

276

Flavell, C. A., Gordon, S., & Watt, K. (2017). Intra-examiner reliability of lumbar

spine and neuro-dynamic flexibility measurements in an older and overweight

healthy asymptomatic population. *Journal of Back and Musculoskeletal*

Rehabilitation, *30*(1), 79-84. DOI 201610.3233/BMR-160717

This administrative form
has been removed

This administrative form
has been removed

278

Intra-examiner reliability of lumbar spine and neuro-dynamic flexibility measurements in an older and overweight healthy asymptomatic population

Carol Ann Flavell[a,*], Susan Gordon[a] and Kerrianne Watt[b]
[a]College of Healthcare Sciences, James Cook University, Townsville, Qld, Australia
[b]College of Public Health, Medical and Veterinary Sciences, James Cook University, Douglas Campus, Townsville, Qld, Australia

Abstract.

BACKGROUND: Measurement of lumbar spine movement and neuro-dynamic flexibility is fundamental to clinical examination and forms part of most systems or protocols used to classify patients with low back pain. However, the increased age and body mass index reported in the chronic low back pain subgroup, may compromise the reliability of these measurements. Specifically, this relates to greater soft tissue excursion relative to the underlying bony landmarks used for these clinical tests.

OBJECTIVE: The objective of this research was to determine the intra-examiner reliability for six lumbar spine and two neuro-dynamic examination tests in older and overweight individuals selected to represent a similar age and body mass index of a chronic low back pain population.

METHODS: Nineteen volunteers (56.00 $\pm$ 7.62 years) performed sets of eight lumbar spine examination tests. Five repetitions of each set of tests were conducted with rest periods between sets. One examiner measured lumbar flexion, extension, right and left lateral flexion and rotation using a tape measure. A goniometer was used to measure Slump and Passive Straight Leg Raise (PSLR) test.

RESULTS: Intra-class correlation coefficients (ICC_{intra}) were calculated to evaluate the reliability of measurement for each test. The eight tests showed ICC ranges between 'Large' and 'Nearly Perfect' (0.68–0.99). Measurement of lateral flexion and rotation had the highest reliability and extension the lowest.

CONCLUSIONS: This study was conducted on an asymptomatic older, overweight population and the ICC results support the suitability of these methods and tools for measurement in a clinical setting for this population demographic. Measurement reliability in the study population was comparable with previous research in populations of contrasting demographics. Accordingly, further investigation specifically in a chronic low back pain population is indicated as the paucity of previous literature for some of the movement tests disallowed comparisons, and the reliability reported in this study may not be directly transferable to a chronic low back pain population.

Keywords: Adult, spine, movement, measurement, intra-examiner reliability

1. Introduction

*Corresponding author: Carol Flavell, College of Healthcare Sciences, James Cook University, Townsville, Qld, Australia. Tel.: +61 7 4781 6472; Fax: +61 7 4781 6868; E-mail: carol.flavell@jcu.edu.au.

Health professionals such as physiotherapists, use spinal movement and neuro-dynamic flexibility tests as measures of intervention effect. Such movement measurements constitute part of an overall assessment pro-

cess for patients with low back pain, but according to previous research, reliability of lumbar spine movement tests is dependent on many factors, including the type of population being measured [1]. There is potential for increased measurement error due to greater soft tissue excursion relative to underlying bony landmarks in older people [2], and excessive adipose tissue relative to underlying bony landmarks in those with a high Body Mass Index (BMI) [3]. Moreover, clinicians should consider the evidence that suggests a relationship between obesity and low back pain [4,5] and the importance of being able to reliably assess spinal movement in obese people with chronic low back pain (CLBP).

Measurement with tape measures and goniometry is used clinically by many health professionals, particularly physiotherapists. Clinicians and researchers use these simple tools to measure spinal movement and neuro-dynamic flexibility, predominantly, to provide an initial measure and then to evaluate change via repeated measures following an intervention. The degree of error in these measurements when repeated over time should be minimal otherwise the results will not represent true change. Reliability reports for lumbar spine and neuro-dynamic movement tests have varied. Beattie et al. [6] evaluated the Modified Schöbers Method for lumbar extension measurement, and Leard et al. [7] the Schöbers Method [8] for lumbar flexion measurement. These studies reported good intra-examiner reliability for lumbar extension and flexion measurement respectively. However, Frost et al. [9] reported good reliability only for lumbar flexion and lateral flexion, and poor reliability for extension and rotation measurement. However, in comparison to Beattie et al. [6], these researchers used an alternative method, which resulted in combined thoraco-lumbar extension measurement, not specific to measurement of lumbar movement. In addition, despite using different measurement methods, both Frost et al. [9] using a tape measure to plinth method, and Leard et al. [7] using inclinometry, reported poor reliability for hip flexion during passive straight leg raising (PSLR), and one study reported good reliability when measuring knee extension during slump test [10].

Various methods for measuring lumbar range of motion (ROM) with a tape measure have been described [8,11–13]. Specifically, Hyytiäinen et al. [12] showed that the Modified Schöber Method [6] to assess lumbar sagittal movements in normal subjects had greater accuracy and higher intra-examiner reliability for measuring lumbar flexion than the Schöbers

Method [8]. Further, the Modified Schöber Method has been validated by good correlation with radiographic measures [13]. Also greater reliability has been reported when using the Modified Schöbers Method for assessment of flexion and extension when compared to a two-inclinometer method [11]. Other common measurements of lumbar spine movement have good intra-examiner reliability, specifically, the finger to floor method to measure lateral flexion [9]. In contrast, use of goniometry for hip joint measurement during neuro-dynamic testing during the passive straight leg raise [7] has been reported to have low intra-examiner reliability. However, measurement of knee extension with electro-goniometry has reported good reliability during the Slump test [10]. Nevertheless, electro-goniometers are more difficult to use than simple tools such as the universal goniometer and may not be readily available in the clinical setting. Tape measure and universal goniometers remain simple, user friendly and frequently available clinical tools for joint measurement. While the results of previous reliability studies are useful, all of them were conducted with participants younger than 50 years and with normal BMI (20–24) [6,7,9,10,12]. Therefore, the results are not specific to an older or overweight population,

Using standard protocols and equipment, this pilot study investigated the intra-examiner reliability of lumbar spine range of motion and neuro-dynamic flexibility in a healthy asymptomatic population of overweight, older adults, selected to represent the demography of the CLBP population.

2. Materials and methods

2.1. Study design

Test-retest intra-examiner single group reliability study.

2.2. Ethics

All participants received a verbal and written description of the study. Written consent was obtained prior to data collection, and the study was conducted with approval of the institutional Human Ethics Research committee (H4547).

2.3. Recruitment

Recruitment was by email, word of mouth, 'flyers' distributed within the local community, and attendance at a community meeting.

2.4. Participants

Participants aged 50 years or over with a BMI of 24 or greater were recruited to the study. Participants were excluded in the presence of any musculo-skeletal or medical condition that would preclude them from safely performing any part of the testing protocol.

2.5. Experimental procedures

2.5.1. Experimental equipment and measurements

Standard protocols were used to measure height with a stadiometer, weight with weighing scales [14], lumbar spine flexion, extension, lateral flexion and rotation with a tape measure, and hip and knee flexion with a goniometer [15].

Six lumbar spine movements were measured; flexion, extension, right and left lateral flexion, and right and left rotation. In addition, to assess neuro-dynamic flexibility range of hip flexion during passive straight leg raise (PSLR) [16] and knee extension during the slump test [17] were measured.

2.5.2. Examiner

The examiner, a physiotherapist with 27 years clinical experience, conducted all examination tests and measurements. An assistant recorder, experienced in reading a goniometer and standard tape measure, recorded each measurement.

2.5.3. Participants

Participants consented to remove upper body outer garments, shoes, and wore shorts during all procedures. Participants were advised and instructed on the study procedures and given the opportunity to warm up with simple active movements prior to testing. In order to avoid any possible risk of injury or discomfort during testing, participants were instructed to perform the test movements to their maximum ability but not 'over stretch' into the movements.

2.5.4. Examination tests

Participants completed eight examination tests using standard measurement protocols [15]; six lumbar spine movement tests measured with the tape measure, and two neuro-dynamic tests measured with the goniometer.

To minimise bias and blind the clinician, the numbers on the face of the measurement devices were covered to the clinician. The assistant was able to view the numbers and recorded each measurement on a stan-
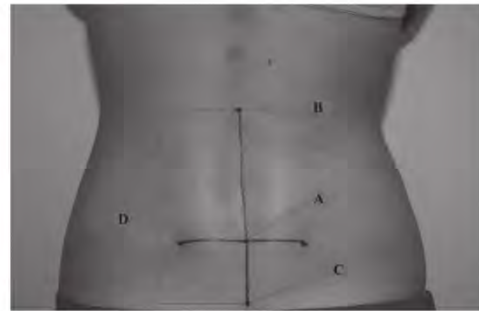

Fig. 1. Illustration of skin landmarks for Modified Schöber test: A. represents the 0 cm mark on the spinous process of S2, identified by a line joining both posterior superior iliac spines. B. represents a mark 10 cm superior to A. C. represents a mark 5 cm inferior to A. D. represents the 15 cm initial measurement distance.

dardised record sheet. All measurements were completed once, then participants rested for fifteen minutes before the movements were repeated and re-measured. This procedure was repeated five times. All tests were recorded during one ninety-minute attendance session.

2.5.5. Lumbar flexion and extension

The skin was marked according to the methods described by Beattie et al. [6] for the Modified Schöbers test as per Fig. 1. For lumbar flexion, participants were asked to bend forwards from a standing position, sliding both hands along the anterior aspect of both thighs towards the floor. For lumbar extension, with hands resting on iliac crests, participants were asked to bend backwards from a standing position. For both tests, once the participant reached their maximum ROM, the tape measure was then placed along a line between landmarks C and B and the distance measured (Fig. 1). The resultant difference between initial (15 cm) and final measurement was calculated to determine ROM.

2.5.6. Right and left trunk lateral flexion

Participants were asked to laterally flex to the right and reach towards the floor while sliding their ipsilateral hand with fingers extended along the lateral aspect of their lower limb. Upon reaching their maximum ROM the examiner used a standard tape measure to record the distance between the tip of the third digit and a point directly inferior to this on the floor. The procedure was then conducted to the contralateral side.

2.5.7. Right and left trunk rotation

On both the left and right side the skin overlying the midpoint of the lateral border of the acromion and the greater trochanter was identified and marked. With arms crossed in front of their chest, participants were asked to turn to the right while keeping their feet fixed on the ground. Upon reaching their maximum ROM, the distance between the markers on the acromion and the greater trochanter was recorded. The procedure was then conducted to the left side.

2.5.8. Unilateral right slump

Participants were seated in a standardised position on the plinth with their hands placed on their anterior thigh and lower legs hanging freely. The middle of the lateral knee joint line on the participant's right side was identified and marked with removable tape. The examiner then conducted a slump test procedure [17] until participants' maximum active right knee extension was achieved. The right knee joint angle was then measured with the goniometer using a standard protocol [15].

2.5.9. Unilateral left PSLR

Participants rested in supine on a treatment plinth without pillows. The participant's left greater trochanter of the femur was identified and marked with removable tape. The examiner then conducted a PSLR procedure [16] by passively raising the left lower limb off the plinth whilst ensuring that the right lower limb remained still with no pelvic tilting. The left knee joint was maintained in an extended position. Maximum PSLR was reached when the participant reported 'maximum' stretch or any discomfort, or the examiner detected firm resistance or pelvic tilting. Participants' left hip flexion was then measured with the goniometer [15].

2.5.10. Statistical analysis

Descriptive statistics including means and standard deviations were calculated for height, weight, age and body mass index (BMI) (Table 1). For each test intra-class correlation coefficients (*ICCs*) with 95% confidence intervals and standard error of measurement [18] were calculated. Statistical analysis was performed using SPSS (version 20) [SPSS inc., Chicago, IL].

Table 1
Participant demographics ($n = 19$)

	Mean	SD	Range
Age (years)	56.00	7.62	(50–81)
Weight (kg)	79.26	14.10	(56.90–105.90)
Height (m)	1.67	0.13	(1.51–1.92)
BMI	28.32	3.58	(24.07–37.08)

Table 2
Intra-examiner reliability

	ICC*	95% CI^	Typical error#
Flexion	0.88	0.78–0.95	0.36
Extension	0.68	0.49–0.84	0.58
Right lateral flexion	0.90	0.82–0.96	0.33
Left lateral flexion	0.94	0.88–0.97	0.27
Right rotation	0.93	0.87–0.97	0.27
Left rotation	0.99	0.98–0.99	0.12
Right slump	0.94	0.89–0.98	0.25
Left PSLR	0.87	0.78–0.95	0.37

Intra Class Correlation Coefficient*; Confidence Interval^; Standard Error of Measurement#, $P < 0.05$**.

3. Results

Nineteen volunteers (8 male, 11 female), participated in the study (Table 1).

The Intra Class Correlation Coefficients (ICC) were nearly perfect ($ICC \geqslant 0.90$) [18], for right ($ICC = 0.90$) and left ($ICC = 0.94$) lateral flexion, right ($ICC = 0.93$) and left ($ICC = 0.99$) rotation, and slump test ($ICC = 0.94$). Very large correlations were reported for the PSLR ($ICC = 0.87$), and flexion ($ICC = 0.88$) measurements, and a large correlation for extension ($ICC = 0.68$) (Table 2).

4. Discussion

This study was conducted to evaluate the intra-examiner reliability of lumbar spine motion, and joint measurements for evaluating neuro-dynamic flexibility, in an older, overweight population. The study demonstrated that across all tests, the intra-examiner reliability of the measurements was nearly perfect [18] for lumbar spine ROM and neuro-dynamic flexibility. Lateral flexion, rotation and slump tests had the highest level of reliability and extension the lowest.

Older participants with higher BMI were recruited to this study, to replicate the reported association between BMI, age and symptoms of low back pain. However, for ethical reasons, only asymptomatic volunteers were recruited, because the potential for repeated movement testing to provoke a pain response in symptomatic participants was high. Our aim was

to evaluate whether the associated factors of higher BMI and age resulted in different reliability of measurement compared to previous research. It is unlikely that intra-examiner reliability would have been different if symptomatic participants had been recruited for this study. Furthermore, consistency of clinician measurement would remain unchanged, regardless of the potential movement restriction in populations with low back pain.

This study assessed the reliability of the Modified Schöbers Method [6] for lumbar flexion and extension measurement. Results of our study showed that reliability was very high for flexion and extension respectively, and support previous studies conducted using similar methods [8,9,11–13].

Lumbar extension measurement in this study reported high ICC, but the lowest reliability of all the measures. Interestingly, poor reliability of the Modified Schöbers Method for extension has been reported previously, being related to the small amount of excursion available during this movement [9]. In contrast, to this Beattie et al. [6] reported ICCs of ninety and greater when measuring lumbar extension, in both symptom free and symptomatic participants with 'significant limiting LBP'. In summary, the combined findings of the current study, with previous research conducted on younger participants with normal BMI by Frost et al. [9] and Beattie et al. [6] remain conflicted.

Nearly perfect ICCs were observed for the intra-examiner measurements of lateral flexion and rotation in our study. Frost et al. [9] reported good intra-examiner reliability for the third finger to floor method of lateral flexion. Hence, the results of our study for reliability of lateral flexion support the findings of Frost et al. [9]. Conversely, our results did not support the poor reliability for rotation measurement reported by Frost et al. [9].

Repeated measurements of slump test and PSLR for neuro-dynamic flexibility showed nearly perfect and very high correlation respectively. Previous studies have reported poor reliability when measuring hip flexion angle during PSLR [7]. Although few intra-examiner reliability studies have measured joint position during PSLR and slump test, one study identified that pelvic position affected hip range of movement during PSLR [19]. In our study pelvic position was controlled as much as possible, but ultimately reliability of this measurement may have been affected by difficulties in positioning the goniometer while maintaining lower limb position during the test. This was particularly challenging with the individuals in this study

who's higher BMI may have reduced the accuracy of greater trochanter identification. The paucity of previous research into measurement reliability for PSLR prevents comparison.

A limitation to this study was the possible implication of progressive increases in ROM over time due to the visco-elastic properties of the spinal soft tissues. No research has previously reported the optimal time interval for repeated measures of the lumbar spine. In an effort to reduce confounding soft tissue adaptations, based on clinical judgment, the clinician set a minimum fifteen-minute interval between participant measurements. The reliability results of this study support the use of this time interval between measurements to reduce the effect of soft tissue adaptation.

5. Conclusion

This research established levels of intra-examiner reliability of lumbar spine and neuro-dynamic flexibility measurements in an older and overweight healthy asymptomatic population. The findings support the suitability of these methods and tools for measurement in the clinical setting, and suggest that age and increased BMI do not adversely affect repeated measurement reliability. Nevertheless, the paucity of previous literature disallows comparisons with other population demographics for some measures. In particular, future studies should aim to evaluate whether symptomatic chronic low back pain patients have a clinically significant effect on measurement reliability. Particularly the neuro-dynamic measurements of PSLR and Slump test.

Conflict of interest

The authors have no conflict of interest to report.

References

[1] Doriot N, Wang X. Effects of age and gender on maximum voluntary range of motion of the upper body joints. Ergonomics. 2006; 49: 269-281.

[2] Kuo Y-L, Tully EA, Galea MP. Skin movement errors in measurement of sagittal lumbar and hip angles in young and elderly subjects. Gait & Posture. 2008; 27: 264-270.

[3] Broadbent CR, Maxwell WB, Ferrie R, Wilson DJ, Gawne-Cain M, Russell R. Ability of anaesthetists to identify a marked lumbar interspace. Anaesthesia. 2000; 55: 1122-1126.

[4] Cimolin V, Vismara L, Galli M, Zaina F, Negrini S, Capodaglio P. Effects of obesity and chronic low back pain on gait. J of NeuroEng and Rehabil. 2011; 8: 55.

[5] Shiri R, Karppinen J, Leino-Arjas P, Solovieva S, Viikari-Juntura E. The Association Between Obesity and Low Back Pain: A Meta-Analysis. Am J Epidemiol. 2010; 171: 135-154.

[6] Beattie P, Rothstein JM, Lamb RL. Reliability of the attraction method for measuring lumbar spine backward bending. Phys Ther. 1987; 67: 364-9.

[7] Leard JS, Crane BA, Ball KA. Intrarater and Interrater Reliability of 22 Clinical Measures Associated With Lower Quarter Malalignment. J Manipulative Physiol. 2009; 32: 270-276.

[8] Macrae IF, Wright V. Measurement of back movement. Ann Rheum Di. 1969; 28: 584-9.

[9] Frost M, Stuckey S, Smalley LA, Dorman G. Reliability of measuring trunk motions in centimeters. Phys Ther. 1982; 62: 1431-7.

[10] Tucker N, Reid D, McNair P. Reliability and Measurement Error of Active Knee Extension Range of Motion in a Modified Slump Test Position: A Pilot Study. JMMT. 2007; 15: E85-E91.

[11] Gill K, Krag MH, Johnson GB, Haugh LD, Pope MH. Repeatability of Four Clinical Methods for Assessment of Lumbar Spinal Motion. Spine. 1988; 13: 50-53.

[12] Hyytiäinen K, Salminen JJ, Suvitie T, Wickström G, Pentti J. Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. Scand J Rehabil Med. 1991; 23: 3-10.

[13] Portek I, Pearcy MJ, Reader GP, Mowat AG. Correlation between radiographic and clinical measurement of lumbar spine movement. Br J Rheumatol. 1983; 22: 197-205.

[14] Norton K, Olds T. Anthropometrica: a textbook of body measurement for sports and health courses, New Delhi, India: CBS Publishers & Distributers PVt. Ltd; 2006.

[15] Clarkson HM. Joint motion and function assessment: A research based practical guide, Philadelphia, USA: Lippincott Williams and Wilkins; 2005.

[16] Goddard MD, Reid JD. Movements induced by straight leg raising in the lumbo-sacral roots, nerves and plexus, and in the intrapelvic section of the sciatic nerve. Journal of J Neurol Neurosurg Psychiatry. 1965; 28: 12-18.

[17] Maitland GO. The slump test: Examination and treatment. Aust J Physiother. 1986; 31: 215-19.

[18] Hopkins WG. Reliability from consecutive pairs of trials. A new view of statistic. 2011. Internet Society for Sport Science 2000, viewed 24 November 2013, http://sportsci.org/resource/stats/.

[19] Cameron DM, Bohannon RW, Owen SV. Influence of Hip Position on Measurements of the Straight Leg Raise Test. J Orthop Sports Phys Ther. 1994; 19: 168-172.

284

Flavell, C. A., Gordon, S., & Marshman, L. (2016). Classification characteristics of a chronic low back pain population using a combined McKenzie and patho-anatomical assessment. *Manual therapy*, *26*, 201-207. DOI: 10.1016/j.math.2016.10.002

This administrative form
has been removed

This administrative form
has been removed

286

This administrative form
has been removed

This administrative form
has been removed

This administrative form
has been removed

290

Original article

Classification characteristics of a chronic low back pain population using a combined McKenzie and patho-anatomical assessment

CrossMark

Carol Ann Flavell [a, *], Susan Gordon [b], Laurence Marshman [c]

[a] College of Healthcare Sciences, James Cook University, Douglas Campus, Townsville, Qld, 4811, Australia
[b] School of Health Sciences, Rehabilitation and Aged Care, Flinders University, Adelaide, SA, 5000, Australia
[c] The Townsville Hospital, Angus Smith Drive, Townsville, Qld, 4811, Australia

A R T I C L E   I N F O

A B S T R A C T

Background: Physiotherapists use musculo-skeletal classification systems for patient assessment. Since its early development, the McKenzie lumbar spine assessment (MK) has been incorporated into examination algorithms and combined with a series of patho-anatomical diagnostic tests. No previous studies have used a MK and a combined examination (MK-C) to provide a detailed profile of patients, report and compare the classification characteristics of a chronic low back pain (CLBP) population.
Objective: To report the classification characteristics of a CLBP population using MK and MK-C examinations, and conduct inter-classification comparison of the MK-C for demographics, the Oswestry Disability Index (ODI), Roland Morris Disability Index (RM), Modified Somatic Perceptions Questionnaire (MSPQ), symptom duration and intensity.
Method: A prospective cross-sectional study conducted in a spinal clinic by a MK trained physiotherapist.
Results: Results were obtained in 150 patients. Using MK, 31% ($n = 47$) of participants were classified as inconclusive. Following MK-C only 6% of participants remained inconclusive ($n = 9$). The most frequent MK-C classification was facet joint syndrome (FJS) (49%). Participants with FJS were significantly older than those classified as discogenic ($p < 0.001$; CI 3.96 19.74), or mixed ($p < 0.001$; CI 5.98 36.41). Participants classified as discogenic had significantly higher RM ($p = 0.022$) and MSPQ ($p = 0.005$) scores than FJS.
Conclusion: Results indicated that 94% of CLBP patients could be classified using a MK-C. The most common presentation in CLBP was facet joint syndrome. Age, RM and MSPQ appeared to be distinguishing characteristics of this population. Future studies should be conducted to establish the validity and reliability of the MK-C.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

A variety of musculo-skeletal classification systems are currently used by physiotherapists for patient assessment. The type of system used varies according to the target anatomical region and the physiotherapist's training and personal preference (McKenzie, 1981; Petersen et al., 2003; Sahrmann et al., 2003; O'Sullivan, 2005; Hill et al., 2008). Each classification follows a structured process, involving an initial history and physical examination, culminating in a final classification, or patho-anatomical diagnosis.

These systems enable planning and implementation of focused treatment approaches.

Embedded in contemporary physiotherapy practice is the McKenzie Classification System, which utilises a movement based examination process (McKenzie, 1981). The McKenzie Institute Lumbar Spine Assessment (McKenzie Institute International, 2006) (MK), is an examination used to classify patients with spinal symptoms into three main syndromes, derangement, dysfunction or postural (McKenzie and May 2003). Patients who do not fit these syndromes are classified as 'other'. Since it's early development this assessment has been incorporated into examination algorithms that have included a series of patho-anatomical diagnostic tests for facet syndrome (FJS), sacro-iliac joint syndrome (SIJS), clinical instability (CI), spinal stenosis, and myofascial presentations (Laslett and van Wijmen, 1999; Petersen et al., 2003; Eirikstoft and

* Corresponding author.
E-mail addresses: carol.flavell@jcu.edu.au (C.A. Flavell), sue.gordon@flinders.edu.au (S. Gordon), l.a.g.marshman@btinternet.com (L. Marshman).

Kongsted, 2014). Primarily, the researchers who proposed these classification algorithms did so to detect subgroups of low back pain (LBP) (Laslett and van Wijmen, 1999), where treatment application would be based on the system of diagnosis and not therapists' particular system preferences (Petersen et al., 2003). Although test validity, and subsequent diagnostic accuracy of these algorithms has yet to be established, they provided classification beyond movement, functional or motor control bases and encompassed a biopsycho-social approach to LBP management. Furthermore, their clinical utility to rule out sinister pathology may transcend their validity to diagnose a specific condition (Cook and Hegedus, 2011). A combined examination which consists of the MK immediately followed by a series of clustered patho-anatomical clinical tests may also provide a detailed profile of patients and their LBP characteristics.

Studies which report and compare demographic, functional and symptom characteristics are limited (Eirikstoft and Kongsted, 2014). Furthermore, no studies have reported and compared the classification characteristics of a chronic low back pain (CLBP) population using a MK and a combined examination (MK-C), despite evidence that some of these characteristics differ in this LBP subgroup. For example, a higher proportion of female patients (DePalma et al., 2012; Viniol et al., 2013), and increased body mass index (BMI) has been associated with CLBP (Heuch et al., 2010; Cimolin et al., 2011). Evidence also suggests that increased age raises the probability of FJS and SIJS (DePalma et al., 2011), and modified somatic pain perceptions are often a feature of chronic pain conditions (Ardic, 2002). Furthermore, increased peripheral symptom distribution has been associated with the longer duration and increased intensity of symptoms reported in CLBP (Prins et al., 2013).

Accordingly, the objectives of this study were to report the classification characteristics of a CLBP population using MK and MK-C examinations, and conduct inter-classification comparison of the MK-C for demographics, the Oswestry Disability Index (ODI), Roland Morris Disability Index (RM), Modified Somatic Perceptions Questionnaire (MSPQ), symptom duration and intensity.

2. Materials and methods

2.1. Study design

A prospective cross-sectional study registered with the Australian New Zealand Clinical Trials Registry (ANZCTR) on 6/03/2013 (ACTRN: 12613000267752) and conducted in a spinal clinic of an outer regional hospital by a MK trained physiotherapist, between July 2012 and March 2014. Patients who attended the clinic resided in outer regional, rural and remote Australian communities and were referred from both primary and secondary health care sources.

2.2. Ethics

Study approval was obtained from the human research ethics committee of James Cook University and the Townsville Hospital (JCUH4387/HREC10QTHS53). Participants received a detailed explanation of the study and provided written consent prior to commencement of the study.

2.3. Participants and examiner

A convenience sample of consecutive patients who attended the weekly specialist CLBP clinic were recruited to the study. A McKenzie credentialed physiotherapist, with more than 15 years of experience using the MK, and almost 30 years of experience in LBP

examination, screened volunteers for inclusion and exclusion criteria, and conducted all examination procedures. The physiotherapist was blinded to medical records and imaging results for the purposes of this study.

2.4. Inclusion criteria

Participants 18 years of age or older, who presented with CLBP, defined as pain between the level of T12 and the buttock crease, with or without associated lower limb symptoms (Henschke et al., 2006) that had persisted for longer than twelve weeks (Reneman et al., 2006) and was adversely influencing functional status, were included.

2.5. Exclusion criteria

Exclusion criteria were previous surgery to the lumbo-sacral spine, inability to tolerate the physical examination, presence of medical "red flags" indicative of potentially serious medical conditions, pregnancy, inability to communicate, progressive neurological disturbance, current litigation, insurance or other compensation claims, previous lumbar spine infection, tumour, fracture or osteoporosis, and medical conditions that result in CLBP, such as fibromyalgia, or ankylosing spondylitis.

2.6. Demographic and CLBP characteristics data

Participant age (yr), gender, weight (kg), height (m), and symptom duration were recorded. A series of self-reported questionnaires were completed prior to examination which included an 11 point visual analogue scale (VAS) (Huskisson, 1974), the Oswestry Disability Index (ODI) (Fairbank et al., 1980; Fairbank and Pynsent, 2000), the Roland Morris Disability Index (RM) (Roland and Morris, 1983) and the Modified Somatic Perceptions Questionnaire (MSPQ) (Main, 1983) to assess pain intensity, disability and somatic pain perception respectively. The VAS was chosen because it has good reliability and higher sensitivity for reporting pain than other measures (Huskisson, 1974; Hawker et al., 2011). The ODI and RM were chosen because they have high sensitivity for reporting disability (Leclaire et al., 1997), validity and good reliability (Fairbank and Pynsent, 2000; Davidson and Keating, 2005; Fritz and Irrgang, 2001). The MSPQ is relevant and appropriate for use in a CLBP study as it was devised and evaluated specifically for CLBP and an MSPQ score > 13 is one of the clinical predictors of lumbar FJS (Laslett et al. 2006a). In addition, it is unlike most other psychological tests for distress which are not sensitive in CLBP (Main, 1983; Main et al., 1992).

2.7. Examination procedure

In a private examination room the physiotherapist completed the MK-C which comprised a standard MK (MII, 2006) followed by a series of clinical tests to identify, secondary or co-existing patho-anatomical sources of symptoms. Each examination was conducted in a consistent and sequential manner (Fig. 1). Findings of all assessment items were documented on a McKenzie lumbar spine assessment form (MII, 2006), plus a supplementary purpose designed assessment sheet for all components of the MK-C. Following completion of the MK, a classification of derangement syndrome, dysfunction syndrome, postural syndrome, mechanically inconclusive or 'other', was assigned according to the following history and clinical examination responses.

Participants were classified with a derangement syndrome if; History included the following:
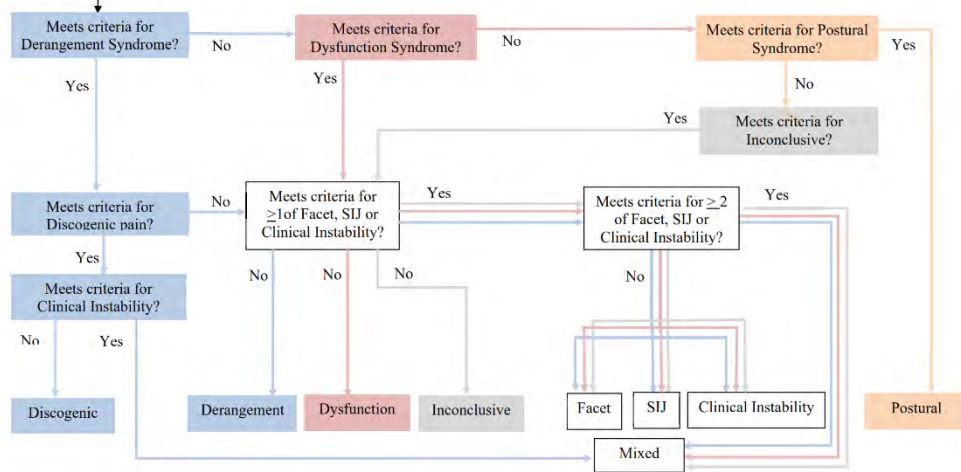
History and Physical Examination

Fig. 1. MK-C examination process.

- Local and/or referred pain which was constant or intermittent.
- Symptoms were episodic in nature, variable over time, with an acute or gradual onset.
- Symptom severity and distribution were aggravated or improved by certain postures or movements.

Physical examination included the following:

- Reduced or 'blocked' range of motion (ROM).
- Repeated movement testing centralised or peripheralised pain and increased or decreased ROM. Centralisation phenomenon has been described elsewhere (May and Aina, 2012).

Participants with derangement syndrome were sub-classified as central symmetrical, asymmetrical above knee, or asymmetrical below knee (McKenzie and May 2003).

Participants were classified as postural syndrome if; History included the following:

- Younger age
- Sedentary lifestyle.
- Time dependent onset of intermittent, local symptoms, due to prolonged postural loading of healthy tissues.
- Symptoms commonly produced with slumped sitting, never produced by movement or activity.

Physical examination included the following:

- Poor spinal posture.
- Postural correction abolished symptoms.
- Static testing produced or abolished pain.
- ROM unaffected.
- No effect with repeated movement testing.

Participants were classified with dysfunction syndrome if; History included the following:

- Previous derangement syndrome or traumatic injury, persistent poor posture, or spinal degeneration.
- Intermittent symptoms that abated and did not persist once loading stopped.
- Local pain, unless symptoms are referred into the lower limb due to adherence of at least one lumbo-sacral nerve root (ANR) (McKenzie and May 2003).

Physical examination included the following:

- Pain production was consistent with a specific direction and ROM.
- ROM reduced in one or more anatomical planes.
- Specific repeated movements produced but never worsened symptoms.

Participants were classified as inconclusive if symptoms were of unknown lumbar joint pathology, affected by movement or posture, but with variable loading response where no one syndrome was predominant on first assessment (McKenzie and May 2003).

Participants classified as 'other' included those with symptoms unrelated to the lumbar spine or sacro-iliac joint, suspected chronic pain states where mechanical responses to assessment were masked by psychosocial or neurophysiological factors, and when no one previously described syndrome was present (McKenzie and May 2003). McKenzie & May (2003) described several presentations as 'other', including suspected serious spinal pathology and the presence of 'red flags'. However, as these presentations were exclusion criteria for this study, 'other' was not represented.

Immediately following the MK the physiotherapist conducted the additional MK-C clinical tests. The MK-C consisted of physical examination processes (Hengeveld et al., 2014), and clustered patho-anatomical clinical tests (Kotilainen and Valtonen, 1993; Laslett et al., 2003; Abbott et al., 2005; Fritz et al., 2005; Laslett et al., 2005a, 2005b, Laslett et al., 2006a; Wilde et al., 2007;

Table 1
Clinical indicators for facet joint syndrome[a] (Hengeveld et al., 2014; Wilde et al., 2007; Laslett et al., 2006a).

	Clinical indicator	MK-C examination component
1	Absence of centralisation.	MK Examination
2	Localised unilateral low back pain/Paraspinal pain.	Body chart
3	Unilateral pressure over facet joint at suspected level replicates or aggravates pain, +/− unilateral muscle spasm.	PAIVM of lumbar spine
4	Lack of radicular features.	Body chart
5	Pain eased in flexion or sitting/Pain not worse with forward flexion.	MK Examination
6	Pain if referred to the leg is above the knee.	Body chart
7	Local unilateral passive movement shows reduced range of motion or increased stiffness ipsilateral to the pain.	PAIVM of lumbar spine
8	Painful lumbar spine extension.	MK Examination
9	Painful lumbar spine extension, lateral flexion or rotation to the ipsilateral side/+ve extension rotation test.	Combined movement testing

[a] Minimum criteria required 1 plus 2,5,6,& 9; MK, McKenzie assessment; MK−C, McKenzie and combined patho-anatomical examination; PAIVM, Passive accessory intervertebral movement.

Laslett, 2008) conducted in the following order:

- Combined lumbar movements extension, ipsilateral rotation and lateral flexion conducted bilaterally (Hengeveld et al., 2014).
- Central and unilateral passive accessory intervertebral movement (PAIVM) T10 to L5/S1 (Hengeveld et al., 2014).
- Nine clinical indicators of FJS (Laslett et al., 2006a; Wilde et al., 2007) (Table 1).
- Six clinical indicators of SIJ (Laslett et al., 2003; Laslett et al., 2005a, 2005b) (Table 2).
- Five clinical indicators of lumbar spine instability (Kotilainen and Valtonen 1993; Abbott et al., 2005; Fritz et al., 2005) (Table 3).

Three of the 12 clinical indicators for FJS identified by Wilde et al. (2007) were not included in the algorithm because they were either not relevant to a physical examination, or were imaging results which for this study the physiotherapist was blinded. Sacro-iliac joint provocation tests were conducted in a consistent order of distraction, compression, thigh thrust, Gaeslen's and sacral thrust according to methods described previously (Laslett et al., 2003; Laslett et al., 2005a, 2005b). On completion of the MK-C participants were assigned to one of nine final classifications: derangement syndrome, postural syndrome, dysfunction syndrome, inconclusive, discogenic, FJS, or SIJS, CI, and mixed, according to operational definitions previously reported (Kotilainen and Valtonen, 1993; Laslett et al., 2003; McKenzie and May 2003; Abbott et al., 2005; Fritz et al., 2005; Laslett et al., 2005a, 2005b; Laslett et al., 2006a; Wilde et al., 2007; Laslett, 2008) and the following history and clinical examination responses.

2.7.1. Discogenic
Participants classified as derangement syndrome with the MK, who presented with centralisation combined with either or both reduced lumbar extension, or a feeling of vulnerability on flexion, and demonstrated no indicators for CI with the MK-C. This test cluster demonstrated validity to predict discogenic pathology when tested in the presence of persistent LBP, against a reference

standard provocation discography. A positive likelihood ratio (+LR) (95% CI) of 6.7 (0.95, 50.0) and a negative likelihood ratio (-LR) of 0.73 (0.63, 0.97) was reported (Laslett et al., 2005a, 2005b; Laslett et al., 2006b).

2.7.2. Derangement
Participants classified as derangement syndrome with the MK, who demonstrated no indicators of discogenic pain, FJS, SIJS or CI with the MK-C.

2.7.3. Dysfunction
Participants classified as dysfunction syndrome with the MK, who demonstrated no indicators of FJS, SIJS or CI with the MK-C.

2.7.4. Inconclusive
Participants classified as inconclusive with the MK, who demonstrated no indicators of FJS, SIJS or CI with the MK-C.

2.7.5. FJS
Participants classified as dysfunction syndrome, inconclusive, or derangement syndrome with the MK, who demonstrated no indicators of discogenic pain, had a MSPQ score >13 and tested positive to elements 2, 5, 6, and 9 of the MK-C FJS cluster (Table 1). Elements of this test cluster have demonstrated validity to rule out facet joint pathology, when tested against a reference standard of facet joint injections, with a +LR 7.6 (4.5, 13.7) and a -LR 0.0 (0.0, 0.35) (Laslett et al., 2006a).

2.7.6. SIJ syndrome
Participants classified as dysfunction syndrome, inconclusive, and derangement syndrome with the MK, who demonstrated no indicators of discogenic pain, and whose symptoms were elicited, or increased with three or more of the SIJ provocation tests(Laslett et al., 2003; Laslett et al., 2005a, 2005b; Laslett, 2008) (Table 2). The validity of this clinical prediction rule to detect SIJS has been reported using SIJ diagnostic injections as a reference standard, with a +LR 6.97 (2.39, 20) and a -LR 0.10 (0.02, 0.68) (Laslett, 2008).

Table 2
Clinical indicators for sacro-iliac joint syndrome (Laslett et al., 2003; Laslett et al., 2005a, 2005b; Laslett, 2008).

	Clinical indicator[a]	MK-C examination component immediately following tests for FJS
1	Absence of centralisation	MK Examination
2	+ve Distraction test	SIJ provocation tests
3	+ve Compression test	
4	+ve Thigh thrust	
5	+ve Gaenslen test	
6	+ve Sacral thrust	

FJS, facet joint syndrome; MK, McKenzie assessment; MK−C, McKenzie and combined patho-anatomical examination; SIJ, sacro-iliac joint.
[a] Minimum criteria required 1 plus >3 SIJ provocation tests.

2.7.7. Clinical instability

This classification was assigned in participants with dysfunction syndrome, inconclusive, and derangement syndrome on the MK, who demonstrated no indicators of discogenic pain, and tested positive for all five MK-C CI indicator tests (Table 3). The MK-C CI indicators were clinically deduced, as tests for CI have previously demonstrated equivocal validity. However, small to moderate + LRs have been reported for Indicator 3, +LR 2.4 (0.9; 6.4) and -LR 0.7 (0.4; 1.0), indicator 4, +LR 1.4 (0.8,2.5) and -LR 0.7 (0.4 .1.2), and indicator 5, +LR 1.1 (0.7,1.8) and -LR 0.9 (0.5 .1.5) (Alqarni et al. 2011). Indicator 1 and 2 were assessed during the MK, and indicator 3, 4, and 5 using methods previously described (Kotilainen and Valtonen, 1993; Abbott et al., 2005; Fritz et al., 2005).

2.7.8. Mixed

Classified according to two criteria;

- Participants classified as derangement syndrome on the MK, with indicators of discogenic pain, and positive responses for all indicators of CI on the MK-C.
- Participants classified as derangement syndrome on the MK, without indicators of discogenic pain, dysfunction syndrome or inconclusive, and demonstrated positive responses for ≥2 test clusters for FJS, SIJS or CI, where no one of these classifications predominated on the MK-C.

2.7.9. Postural syndrome

Based on McKenzie's definition of postural syndrome (McKenzie and May 2003), further examination for patho-anatomical sources of symptoms was not indicated, and following the MK-C examination the final classification for these participants remained as postural syndrome.

2.8. Data management

Baseline ODI, RM and MSPQ were converted to a percentage of raw score, and BMI as a measure of participant's relative size, was calculated from height and weight measurements as kg/m² (Dahl et al., 2013). All data was imported to SPSS statistical analysis software, version 22 (SPSS inc., Chicago, IL, USA).

Descriptive statistics were generated and numerical variables were analysed for normality of distribution and reported using mean and standard deviation (SD), or median with inter-quartile range (IQR). Categorical variables were reported as proportions (%). Bivariate statistical tests were conducted as appropriate to assess for significant differences ($p < 0.05$) in demographic and CLBP characteristics, between diagnostic classification categories. The dysfunction syndrome and dysfunction ANR.

classification were pooled for statistical analysis purposes. All MK and MK-C data was complete and no indeterminate results

existed. Classification categories with less than five participants were reported in this study but no statistical analysis was conducted as numbers were considered too low for statistical inference.

3. Results

Between July 2012 and March 2014, all patients who attended the CLBP clinic ($n = 316$) were invited to the study. The study recruited 62% ($n = 197$) of invited participants. Following application of inclusion and exclusion criteria 76% ($n = 150$) of the 197 volunteers were admitted to the study. Reasons for exclusion included, previous surgery to the lumbo-sacral spine ($n = 9$), current litigation, insurance or other compensation claims ($n = 4$), patients with non-lumbar spinal pain ($n = 6$), ankylosing spondylitis or systemic lupus erythematosus ($n = 2$), inability to tolerate the physical examination due to poor standing balance or other limiting co-morbidities ($n = 26$).

3.1. Demographic characteristics

Demographic and examination characteristics are detailed in Table 4. Males represented 35% and females 65% of participants. Males reported significantly lower.

MSPQ ($p = 0.035$; CI -10.45−−0.4) and ODI ($p = 0.032$; CI -10.28−−0.46) scores than females. No other significant differences existed between genders for any descriptive factor.

3.2. Classification profile of MK

Participant examination most frequently resulted in a McKenzie classification of dysfunction syndrome (36%), derangement syndrome (32%), or mechanically inconclusive (31%), and least frequently postural syndrome (1%).

3.3. Classification profile of MK-C

Table 5 summarises the results for all MK-C classifications. Participant examination most frequently resulted in a MK-C classification of FJS (49%) and least frequently SIJS (0.7%). Of all participants with a McKenzie classification of derangement syndrome, eight were negative for the intervertebral disc as the source of symptoms. Of these cases, four were negative for additional patho-anatomical sources of their symptoms resulting in an MK-C classification of derangement syndrome. The remaining four tested negative for SIJS and CI, but positive for FJS, and therefore were classified as such.

Indicators for discogenic pain were evident in 83% ($n = 39$) of participants with a McKenzie classification of derangement syndrome. Of these cases, 35 were negative for indicators of CI resulting in a final classification of discogenic. The remaining

Table 3
Clinical indicators of clinical instability (Kotilainen and Valtonen, 1993; Abbott et al., 2005; Fritz et al., 2005).

Clinical indicator	MK-C examination component immediately following tests for FJS, & SIJS
1 Reports of recurrent locking, catching, giving way during active movements or apprehension (demonstrates anxiety about the sensation of collapse due to low back pain during movement)	MK movement tests
2 Aberrant motion with trunk ROM eg hand-thigh walking on extension or instability catch sign test (inability to bend forward and return to an erect position due to sudden onset low back pain)	MK movement tests
3 Intervertebral motion testing to determine hypermobility	PAIVM of lumbar spine
4 Prone instability test	SIJ provocation tests
5 Posterior shear test	

FJS, facet joint syndrome; MK, McKenzie assessment; MK−C, McKenzie and combined patho-anatomical examination; PAIVM, Passive accessory intervertebral movement; SIJ, sacro-iliac joint; SIJS, sacro-iliac joint syndrome.

Table 4
Demographic data and CLBP characteristics according to each McKenzie and combined patho-anatomical classification.

	Total $n = 150$	Discogenic $n = 35$	Dysfunction syndrome $n = 12$	Inconclusive $n = 9$	FJS $n = 74$	CI $n = 6$	Mixed $n = 7$
Female %	65	66	58	67	65	50	86
Male %	35	34	42	33	35	50	14
Symptom Duration (months) Median (*IQR*)	36 (52)	36 (54)	35 (94)	60 (141)	36 (47)	75 (171.5)	24 (84)
Age (mean ± *SD*)	56.7 ± 14.32	51.06 ± 14.43[**]	55.92 ± 13.95	57 ± 10.91	62.91 ± 12.48[§**]	47.83 ± 12.29	41.71 ± 8.42[§]
BMI (kg/m^2) (mean ± *SD*)	31.5 ± 7.27	29.84 ± 7.35	34.95 ± 7.52	29.52 ± 6.42	32.05 ± 7.17	28.65 ± 5.17	31.49 ± 6.38
[ᵱ]ODI (mean ± *SD*)	45.2 ± 14.54	49.76 ± 12.33	39.17 ± 13.85	46.81 ± 19.13	43.24 ± 14.38	53.63 ± 12.47	47.46 ± 9.4
[€]RM Median (*IQR*)	54.1 (35.42)	66.67 (29.17)[*]	52.09 (31.28)	50 (41.67)	50 (34.35)[*]	64.58 (51.05)	54.17 (29.17)
[ᵃ]MSPQ Median (*IQR*)	20.5 (20.5)	30.8 (30.8)[*]	15.4 (21.8)	20.5 (37.2)	17.9 (12.8)[*]	37.15 (21.85)	30.08 (28.2)
[ᵗ]VAS Median (*IQR*)	44 (34.5)	59 (40)	42.5 (43.5)	45 (52.5)	41.5 (27)	67 (40.5)	35 (61)

[*]significant at $p < 0.05$[*]; [§] and[**] significant at $p < 0.001$.

participants with a McKenzie classification of derangement syndrome were more complex presentations who tested positive for CI and were classified as mixed ($n = 4$).

Of participants with a McKenzie classification of dysfunction syndrome, 78% ($n = 42$) tested positive for patho-anatomical sources of their symptoms with the MK-C, and were assigned a final classification of FJS ($n = 38$), CI ($n = 3$) and mixed ($n = 1$).

The remaining 22% ($n = 12$) were negative for patho-anatomical sources of their symptoms and were classified as dysfunction syndrome. Following the MK-C only 6% of participants remained inconclusive ($n = 9$). Therefore, 38 of the participants initially classified as inconclusive with the MK ($n = 47$), were classified as FJS ($n = 32$), CI ($n = 3$), SIJS ($n = 1$), or mixed ($n = 2$) following the MK-C examination.

3.4. Characteristics of the MK-C: Inter-classification comparison

Participants classified with FJS were significantly older than those classified as discogenic ($p < 0.001$), or mixed ($p < 0.001$). Participants classified as discogenic had significantly higher RM ($p = 0.022$) and MSPQ ($p = 0.005$) scores than those classified with FJS. In contrast, there were no significant differences in RM and MSPQ scores between any other final classification categories. Similarly, there were no significant differences in ODI ($p = 0.085$), BMI ($p = 0.241$), VAS ($p = 0.631$) or symptom duration ($p = 0.522$) between any MK-C classification categories (Table 4).

4. Discussion

This study found that, by using the MK-C, the most common classification was FJS. Postural syndrome, SIJS, and CI, were classified in a minority of participants. Consistent with previous studies (Razmjou et al., 2000; Clare et al., 2005; Hefford, 2008; Eirikstoft and Kongsted, 2014) this study found less than two percent of participants were classified with CLBP of postural origin.

Table 5
Results of the McKenzie and combined patho-anatomical examination (MK-C).

	$n = 150$	%
Discogenic	35	23.3
Derangement syndrome	4	2.7
Dysfunction syndrome	12	8
Inconclusive	9	6
FJS	74	49.3
SIJS	1	0.7
CI	6	4
Mixed	7	4.7
Postural syndrome	2	1.3

FJS, facet joint syndrome; SIJS, sacro-iliac joint syndrome; CI, clinical instability.

This study identified that successful classification in a CLBP population was increased from 69% using the MK to 94% with the MK-C. McKenzie and May (2003), advocate that for classification using the MK, the patient may need to attend on several consecutive occasions. Indeed, using the MK in this CLBP population resulted in an inconclusive classification at first attendance in 31% of participants. The MK-C appears to be more efficient as it provided a classification at first examination, for the McKenzie syndromes of derangement without discogenic indicators, mechanically inconclusive and dysfunction. This research supports the use of this combined examination method to optimise classification of CLBP patients.

This study provided key findings particularly in relation to participant age, disability and modified somatic pain perceptions between classification categories. To the authors' knowledge the classification characteristics of a CLBP population examined using an MK or MK-C has not been described previously. Patients classified with FJS were significantly older than those classified with discogenic pain or a mixed presentation. These age related classification differences between FJS and symptoms of intervertebral disc origin are explained and supported by previous evidence that the degenerative disc precedes that of the facet joints (Miller et al., 1988; Butler et al., 1990; Fujiwara et al., 1999).

The significantly higher RM and MSPQ score for the discogenic classification compared to FJS is more difficult to explain, noting that the difference in the mean RM scores for the FJS and discogenic classifications was 4/24, which may not be considered clinically significant (Roland and Fairbank, 2000). No significant differences were reported in symptom duration between any classification categories in this cohort suggesting that differences in MSPQ scores between discogenic and FJS were not solely attributable to symptom duration. Further research is required to explore the reasons for higher modified somatic pain perceptions in CLBP patients with chronic discogenic pain.

The findings of this study highlight and support the utility of a MK-C for assessment and classification of CLBP. This enables physiotherapists to utilise diagnostic skills to classify CLBP patients at first visit, whilst maintaining their unique ability to assess according to the fundamentals of movement, symptom provocation and function.

The results of this study are limited as the assessment was conducted by one physiotherapist, and the classification characteristics presented are specific to the cohort tested, and the classification algorithm used for this study. Accordingly, results may not be generalised to other LBP subgroups, other classification systems, and patho-anatomical diagnoses. Nevertheless, in the absence of previous evidence specific to this topic the reported findings are relevant and clinically useful.

5. Conclusion

This study determined that 94% of CLBP patients could be classified using a MK-C. The most common presentation was FJS. The MK-C examination could refine the McKenzie syndromes of derangement without discogenic indicators, mechanically inconclusive and dysfunction syndromes into a classification based on patho-anatomical causes of patient symptoms. Only, age, RM and MSPQ differed amongst classification categories and appeared to be distinguishing characteristics of this population. Future studies should be conducted to establish the diagnostic accuracy, validity and inter-observer reliability of the MK-C.

References

Abbott, J.H., et al., 2005. Lumbar segmental instability: a criterion-related validity study of manual therapy assessment. BMC Musculoskelet. Disord. 6 (56).

Alqarni, A.M., Schneider, A.G., Hendrick, P.A., 2011. Clinical tests to diagnose lumbar segmental instability: a systematic review. J. Orthop. Sports Phys. Ther. 41 (3), 130–140.

Ardic, F., 2002. Psychological dimensions of pain in patients with rheumatoid arthritis, fibromyalgia syndrome, and chronic low back pain. J. Musculoskelet. Pain 10 (4), 19–29.

Butler, D., et al., 1990. Discs degenerate before facets. Spine (Phila. pa. 1976) 15 (2), 111–113.

Cimolin, V., Vismara, L., Galli, M., Zaina, F., Negrini, S., Capodaglio, P., 2011. Effects of obesity and chronic low back pain on gait. J. NeuroEng. Rehabil. 8 (55).

Clare, H.A., Adams, R., Maher, C.G., 2005. Reliability of McKenzie classification of patients with cervical or lumbar pain. J. Manip. Physiol. Ther. 28 (2), 122–127.

Cook, C., Hegedus, E., 2011. Diagnostic utility of clinical tests for spinal dysfunction. Man. Ther. 16 (1), 21–25.

Dahl, A.K., et al., 2013. Body mass index, change in body mass index, and survival in old and very old persons. J. Am. Geriatr. Soc. 61 (4), 512–518.

Davidson, M., Keating, J., 2005. Oswestry disability questionnaire (ODQ). Aust. J. Physiother. 51 (4), 270.

DePalma, M.J., Ketchum, J.M., Saullo, T.R., 2011. What is the source of chronic low back pain and does age play a role? Pain Med. 12, 224–233.

DePalma, M.J., Ketchum, J.M., Saullo, T.R., 2012. Multivariable analyses of the relationships between age, gender, and body mass index and the source of chronic low back pain. Pain Med. 13 (4), 498–506.

Eirikstoft, H., Kongsted, A., 2014. Patient characteristics in low back pain subgroups based on an existing classification system. A descriptive cohort study in chiropractic practice. Man. Ther. 19 (1), 65–71.

Fairbank, J.C.T., Pynsent, P.B., 2000. The Oswestry disability index. Spine 25 (22), 2940–2953.

Fairbank, J.C.T., Couper, J., Davies, J.B., 1980. The Oswestry low back pain questionnaire. Physiotherapy 66, 271–273.

Fritz, J.M., Irrgang, J.J., 2001. A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec low back pain disability scale. Phys. Ther. 81 (2), 776–788.

Fritz, J.M., Piva, S.R., Childs, J.D., 2005. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. Eur. Spine J. 14 (8), 743–750.

Fujiwara, A., et al., 1999. The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study. Eur. Spine J. 8, 396–401.

Hawker, G.A., et al., 2011. Measures of adult pain: visual analog scale for pain (VAS pain), numeric rating scale for pain (NRS pain), McGill pain questionnaire (MPQ), short-form McGill pain questionnaire (SF-MPQ), chronic pain grade scale (CPGS), short Form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP). Arthr. Care Res. 63 (S11), S240–S252.

Hefford, C., 2008. McKenzie classification of mechanical spinal pain: profile of syndromes and directions of preference. Man. Ther. 13 (1), 75–81.

Hengeveld, E., Banks, K., Maitland, G.D., 2014. In: Maitland's Vertebral Manipulation. Volume 1: Management of Neuromusculoskeletal Disorders, eighth ed., vol. I. Churchill Livingstone, Edinburgh.

Henschke, N., et al., 2006. Prognosis of acute low back pain: design of a prospective inception cohort study. BMC Musculoskelet. Disord. 7 (54), 7(54).

Heuch, I., et al., 2010. The impact of body mass index on the prevalence of low back pain the hunt study. Spine (Phila. pa. 1976) 35 (7), 764–768.

Hill, J.C., et al., 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Care & Res. 59 (5), 632–641.

Huskisson, E.C., 1974. Measurement of pain. Lancet 304 (7889), 1127–1131.

Kotilainen, E., Valtonen, S., 1993. Clinical Instability of the lumbar spine after microdiscectomy. Acta Chir. (Wein) 125, 120–126.

Laslett, M., van Wijmen, P., 1999. Low back and referred pain: diagnosis and a proposed new system of classification. N. Z. J. Physiother. 27 (2), 5–14.

Laslett, M., et al., 2003. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. Aust. J. Physiother. 49 (2), 89–97.

Laslett, M., et al., 2005a. Diagnosis of Sacroiliac Joint Pain: validity of individual provocation tests and composites of tests. Man. Ther. 10 (3), 207–218.

Laslett, M., et al., 2005b. Centralization as a predictor of provocation discography results in chronic low back pain, and the influence of disability and distress on diagnostic power. Spine J. 5 (4), 370–380.

Laslett, M., et al., 2006a. Clinical predictors of screening lumbar zygapophysial joint blocks: development of clinical prediction rules. Spine J. 6 (4), 370–379.

Laslett, M., et al., 2006b. Clinical predictors of lumbar provocation discography: a study of clinical predictors of lumbar provocation discography. Eur. Spine J. 15 (10), 1473–1484.

Laslett, M., 2008. Evidence-based diagnosis and treatment of the painful sacroiliac joint. J. Man. Manip. Ther. 16 (3), 142–152.

Leclaire, R., et al., 1997. A cross-sectional study comparing the Oswestry and roland-morris functional disability scales in two populations of patients with low back pain of different levels of severity. Spine 22 (1), 68–71.

Main, C.J., 1983. The modified somatic perception questionnaire (MSPQ). J. Psychosom. Res. 27 (6), 503–514.

Main, C.J., et al., 1992. The Distress and Risk Assessment Method: a simple patient classification to identify distress and evaluate the risk of poor outcome. Spine (Phila. pa. 1976) 17 (1), 42–52.

May, S., Aina, A., 2012. Centralization and directional preference: a systematic review. Man. Ther. 17 (6), 497–506.

McKenzie Institute International, 2006. The McKenzie Institute Lumbar Spine Assessment [cited 2006 3rd December]; Available from: http://mckenziemdt.org/forms/2006MIIAssessmentFormLumbar.pdf.

McKenzie, R.A., 1981. The Lumbar Spine: Mechanical Diagnosis and Therapy. New Zealand: Spinal Publications, Waikanae.

McKenzie, R.A., May, S., 2003. In: The Lumbar Spine Mechanical Diagnosis & Therapy, second ed., vol. 2. New Zealand: Spinal Publications, Waikanae.

Miller, J.A.A., Schmatz, C., Schultz, A.B., 1988. Lumbar disc degeneration: correlation with age, sex, and spine level in 600 autopsy specimens. Spine (Phila. pa. 1976) 13 (2), 173–178.

O'Sullivan, P., 2005. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. Man. Ther. 10 (4), 242–255.

Petersen, T., et al., 2003. Diagnostic classification of non-specific low back pain. A new system integrating patho-anatomic and clinical categories. Physiother. Theory Pract. 19 (4), 213–237.

Prins, M.R., van der Wurff, P., Groen, G.J., 2013. Chronic low back pain patients with accompanying leg pain: the relationship between pain extent and pain intensity, disability and health status. J. Back Musculoskelet. Rehabil. 26 (1), 55–61.

Razmjou, H., Kramer, J.F., Yamada, R., 2000. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. J. Orthop. Sports Phys. Ther. 30 (7), 368–389.

Reneman, M.F., et al., 2006. Symptom increase following a functional capacity evaluation in patients with chronic low back pain: an explorative study of safety. J. Occup. Rehabil. 16 (2), 192–205.

Roland, M.O., Fairbank, J.C.T., 2000. The roland–morris disability questionnaire and the Oswestry disability questionnaire. Spine 25 (24), 3115–3124.

Roland, M.O., Morris, R.W., 1983. A study of the natural history of back pain. Part 1: development of a reliable and sensitive measure of disability in low back pain. Spine 8, 141–144.

Sahrmann, S.A., et al., 2003. Movement system impairment-based categories for low back pain: stage I validation. J. Orthop. Sports Phys. Ther. 33, 126–142.

Viniol, A., et al., 2013. Differences between patients with chronic widespread pain and local chronic low back pain in primary care - a comparative cross-sectional analysis. BMC Musculoskelet. Disord. 14 (1), 1–9.

Wilde, V.E., Ford, J.J., McMeeken, J.J., 2007. Indicators of lumbar zygapophyseal joint pain: survey of an expert panel with the delphi technique. Phys. Ther. 87 (10), 1348–1361.

Flavell, C., Gordon, S., & Marshman, L. (2017). Reply to the letter to the editor regarding 'Classification characteristics of a chronic low back pain population using a combined McKenzie and patho-anatomical assessment'. *Musculoskeletal Science and Practice, 27*, e7. doi:https://doi.org/10.1016/j.msksp.2016.12.012

This administrative form
has been removed

This administrative form
has been removed

This administrative form
has been removed

300

This administrative form
has been removed

302

This administrative form
has been removed

Letter to the Editor

Reply to the letter to the editor regarding 'Classification characteristics of a chronic low back pain population using a combined McKenzie and patho-anatomical assessment'

We would like to address the questions raised in the letter as they are presented. We acknowledge and respect the outstanding work of Robin McKenzie and the work which continues to be conducted by the McKenzie Institute and its members. Combining a McKenzie assessment with patho-anatomical classifications was never meant as a substitute for the McKenzie classifications and subgroupings which is an established classification process. However, we wish to clarify some details.

The syndrome prevalence we report differs to previous studies, but our study sample is not comparable to the populations reported in the previous studies cited (Hefford, 2008; May, 2006; Otero et al., 2014; Werneke et al., 2010, 2016).

We note the following differences:

1. Otero and Bonnet (2014), was a study conducted on a population with symptoms of cervical origin.
2. Hefford (2008), clearly reported that the participants examined by the physiotherapists in her study population were a mixed acute and chronic sample with no participant sample limited to CLBP only. In addition, only 187 of the total 321 participants had symptoms of lumbar origin.
3. None of the other studies represented a 100% CLBP sample.

All participants in our study were recruited from a chronic pain clinic and participants with acute LBP were specifically excluded. Chronic pain conditions are complex and require management with a multi-disciplinary approach (Kamper et al., 2015). CLBP is no exception to this, and is preferentially assessed by a team of health professionals including but not exclusive to, medical specialists, physiotherapists, nurses, social workers and psychologists. In such settings, shared terminology and communication is fundamental to optimal management (Cedraschi et al., 1998). This requires a mutually understood language between disciplines.

Our study does not question the clinical importance of the Mechanical Diagnosis and Therapy approach, nor any other existing method of low back pain classification. Nevertheless, whilst much physiotherapy specific terminology is recognised and accepted within our profession, currently not all is universally understood by other health professionals. This has potential to hinder effective communication with medical specialists, and other health providers. Regardless of debate on the 'flawed' and 'failed' patho-anatomical approach to low back pain management, within a multi-disciplinary environment, related terminology remains accepted and understood between health disciplines.

Therefore, we support the clinical relevance of reporting the characteristics of a CLBP population using a combined examination approach. Particularly, since this has not been conducted previously. There remains much to learn about CLBP, and it is only by evaluating alternative approaches to what is a complex and multi-factorial condition that our knowledge can increase.

Kind regards.

References

Cedraschi, C., Nordin, M., Nachemson, A.L., Vischer, T.L., 1998. 1 health care providers should use a common language in relation to low back pain patients. Bailliere's Clin. Rheumatol. 12 (1), 1–15.

Kamper, S.J., Apeldoorn, A.T., Chiarotto, A., Smeets, R.J.E.M., Ostelo, R.W.J.G., Guzman, J., van Tulder, M.W., 2015. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: cochrane systematic review and meta-analysis. Br. Med. J. 350, 1–11. http://dx.doi.org/10.1136/bmj.h444.

Carol Flavell*
College of Healthcare Sciences, James Cook University, Douglas Campus, Townsville, Qld, 4811, Australia

Susan Gordon
School of Health Sciences, Rehabilitation and Aged Care, Flinders University, Adelaide, SA, 5000, Australia
E-mail address: sue.gordon@flinders.edu.au.

Laurence Marshman
The Townsville Hospital, Angus Smith Drive, Townsville, Qld, 4811, Australia
E-mail address: l.a.g.marshman@btinternet.com.

* Corresponding author.
E-mail address: carol.flavell@jcu.edu.au (C. Flavell).

Appendix 4: Human Research Ethics Approvals

4A: Human Research Ethics Approval James Cook University (JCUH4547)

This administrative form
has been removed

4B: Human Research Ethics Approval James Cook University (JCUH4387), &

The Townsville Hospital (HREC10QTHS53)

This administrative form
has been removed

This administrative form
has been removed

This administrative form
has been removed

4C: Human Research Ethics Approval Mater health services (MHS20150512-07)

This administrative form
has been removed

Appendix 5: Grants and Scholarships

This administrative form
has been removed

This administrative form
has been removed

This administrative form
has been removed

312