Investigating Acupuncture and Manual Therapy for Low Back Pain

Volume 1

Vivienne Claire Dascanio

Doctor of Philosophy

University of York

Health Sciences

July 2018

Abstract

Background

Low Back Pain (LBP) is the principal cause of disability globally (Buchbinder et al. 2018; Hartvigsen et al. 2018). Research and packages of care have strived to reduce levels of LBP, but globally it continues to rise (Foster et al. 2018). There is debate as to the role of acupuncture in LBP care pathways and research on acupuncture is of varying quality and not conclusive (SAR, 2018). Previous research of acupuncture has not effectively used trial design to produce high quality, robust and convincing evidence.

Aims

To investigate if acupuncture and manual therapy are indicated as appropriate treatments for the treatment of LBP and consider if their combination may be viable and effective.

To determine the international LBP recommendations for acupuncture and manual therapy.

To establish which RCT design could best evaluate acupuncture and manual therapy for the treatment of LBP and to trial this in a pilot study.

To ascertain if high-quality research compares and combines acupuncture with manual therapy for LBP.

Methods

To review the international clinical practice guidelines and their approach to acupuncture and manual therapy for LBP. To consider a range of clinical trial designs and establish a preferred design. To conduct a trial investigating acupuncture and manual therapy alone and in combination for LBP, using a cohort study with nested factorial RCT. To conduct a systematic review, comparing acupuncture with manual therapy for LBP.

Results

Further evidence is indicated in the study of acupuncture and manual therapy for LBP. Clinical practice guidelines are inconsistent in their interpretation of evidence and the recommendation of acupuncture. A cohort study with nested factorial RCT is an effective design for recruitment, retention and to evaluate acupuncture and manual therapy for LBP. 97% of participants accepted the interventions offered and 100% of individuals completed the RCT interventions and 100% returned (97% completion of primary outcomes measures) of follow-up questionnaires. Zero attrition was achieved with this pilot study (95% CI 0.0, 6.3). Manual therapy may be superior (-1.4, 95% CI -3.8, 1.0, P=0.24) to usual care, but the results are not statistically significant. Manual therapy appears favourable in an SR and meta-analysis of manual therapy versus acupuncture; the results were limited by the methodological quality of the studies included.

Conclusions

A full-scale definitive trial of acupuncture and manual therapy using a cohort design, with nested factorial RCT is needed.

Acknowledgments

This research and thesis was supported with a three-year PhD studentship from the National Institute for Health Research (NIHR) as part of a programme grant awarded to Professor Hugh MacPherson, for which I am eternally grateful.

Throughout my PhD, I have been privileged to work with some of the best experts. I could not have asked for better teachings or support, and I hope this is reflected in this thesis. A special thank you to Professor Trevor Sheldon for your encouragement, your confidence in me, and supporting me over the finishing line, I am so thankful to you, and to Dr Tracy Lightfoot and the University of York for the opportunity I have received. To Dr Yvonne Birks, Professor David Torgerson, and Professor Hugh MacPherson, thank you for your supervision, support and feedback during my PhD journey. To Caroline Fairhurst for her assistance and support with my statistical analysis. To all the study research participants, the physiotherapists, Doctors and staff at Highthorn Physiotherapy Clinic, Strensall Medical Practice, Jorvik Medical Practice, and the York Trials Unit, thank you for your commitment to this study.

A special thank you goes to, my family and friends for their support and encouragement. Finally, I would like to thank my husband Alberto; we could not have predicted the mammoth journey this PhD would take us on. I thank you for being prepared to relocate to York for me to allow me to fulfil my dream to complete my PhD. Never complaining when I was buried in papers, continually feeding me the most amazing food, and doing more than your share, always believing in me and encouraging me. Thank you, Alberto, for your endless patience and love during all the good times and challenging ones.

This thesis is dedicated to my son Augusto who arrived mid PhD and has been my inspiration to continue and complete my studies. Thank you, Augusto, for your patience with Mummy in your young life. Every day with you is precious, and I hope through my achievement of this thesis I have taught you about dedication, commitment and passion. I look forward to continuing to watch you grow. It has been a journey, a long but enjoyable journey, it has been a real joy to work on this thesis and I am thankful and proud of the person I have become.

Authors Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, nor material which has been accepted for the award of any other degree or diploma at the University or any other institution of higher learning, except where due acknowledgement has been made in the thesis text. The views expressed in the thesis are those of the author and not necessarily the views of others at the University of York, where the research was conducted.

The trial was designed and conducted by myself with input from my supervisors; I was responsible for my research and the data collection for the trial to the six-month time point, within the scope and time frame of this PhD.

Vivienne Claire Dascanio

31st July 2018

Table of Contents

List of Tables:

List of Figures:

List of boxes:

Publications from this thesis

Original Article – Dascanio, VC. Waldock-Goude, A. MacPherson, H. Sheldon, T. (2019) Systematic Review of Manual Therapy Versus Acupuncture for low back pain. To be submitted for publication.

Original Article: Debate Editorial – Dascanio, VC (2015b) Acupuncture in Physiotherapy: a contemporary UK perspective. Acupuncture in Medicine; 33: 442-444

Original Article: Dascanio, V. (2015d) A pragmatic pilot factorial randomized controlled trial of acupuncture versus manual therapy for low back pain nested within an observational cohort study: a trial protocol. Acupuncture in Physiotherapy; 27 (2): 19 – 36

Original Article: - Dascanio, V. Birks, Y. Clark, L. Fairhurst, C. MacPherson, H. Torgerson, D. (2014) Randomized cohort trial was shown to be feasible for evaluating treatments in low back pain. Journal of Clinical Epidemiology, 67, 8, 940-946

Corrigendum to 'Randomized cohort trial was shown to be feasible for evaluating treatments in low back pain'. Journal of Clinical Epidemiology. Volume 67, Issue 8, August 2014, Pages 940-946. Journal of Clinical Epidemiology (2019) Doi: https://doi.org/10.1016/j.jclinepi.2014.04.004

Original Publication: Conference Abstract - Poster Presentation – Dascanio, V. Birks, Y. Torgerson, D. (2011) A pilot factorial randomised cohort trial of manual therapy or acupuncture for low back pain. Trials; 12 (1): A150

Publications associated with this thesis:

Original Publication: Part 1 of Acupuncture in Physiotherapy: The evidence and commissioning resource: Dascanio, V. van Dongen, C. Ireland, CM. Wilson, AJ. (2015e) Acupuncture in Physiotherapy. The Evidence: A summary of evidence for the use of acupuncture in physiotherapy for the benefit of the patient. Publication of: the Acupuncture Association of Chartered Physiotherapists Limited. https://www.aacp.org.uk/assets/ckfinder_library/files/150527%20AACP%20The%20Evidence%20(1).pdf

Original Publication: Part 2 of Acupuncture in Physiotherapy: The evidence and commissioning resource: Dascanio, V. van Dongen, C. Ireland, CM. Wilson, AJ. (2015f) Acupuncture in Physiotherapy: Commissioning Guidance: Guidance on commissioning for AACP members. Publication of: the Acupuncture Association of Chartered Physiotherapists Limited (AACP Ltd). Original Publication. https://www.aacp.org.uk/assets/ckfinder_library/files/150527%20AACP%20Commissioning%20Guidance%20.pdf

Glossary of terms

Acupuncture - Acupuncture is ancient therapy with its history in Chinese medicine. It translates 'to puncture with a needle' (World Health Organisation (WHO), 1991) involving the insertion of fine needles into the body, treating holistically (whole body). In Chinese philosophy illness is considered an imbalance of energy sources in the body, acupuncture strives to recreate balance and harmony (Vera et al. 2018; Chong et al. 2015). In western medicine, acupuncture is considered in physiological terms and is understood to stimulate blood flow, activate nerve activity (Kim et al. 2016; Li et al. 2013; Uchida and Hotta, 2008; Inoue et al. 2005) and stimulate specific areas of the brain to release pain-relieving chemicals (Jin et al. 2016; He et al. 2015; Bradnam 2011; Lewith, White and Pariente, 2005).

Central Sensitization pain – Central sensitization pain occurs due to the dysfunction of neural (nerve) pathway signals of the central nervous system, resulting in a heightened (warning) neural response to painful stimuli (Tatta, 2017).

Cytokines – Are small proteins released by cells in the body, which interact with other cells. They trigger an inflammatory response and respond to infection (Shiel, 2018).

Disability adjusted life years (DALYs) – One lost life year due to disability (WHO, 2018).

Global burden of disease (GBD) – is a comprehensive regional and global research programme of disease burden that assesses mortality and disability of disease, injury and risk factors (WHO, 2018).

Idiopathic – is pain of unknown origin. It is a term used to describe long-term chronic pain without an identifiable cause (Jacques, 2018).

Low back pain (LBP) - is defined as pain, soreness, tension, muscle spasm or stiffness located between the costal margins (bottom of the ribs) and the gluteal folds (top of the legs) presenting with or without leg pain (sciatica) (McIntosh et al, 2011).

Low back pain: Non-specific - is identified as being benign and caused by structures in the back such as joints, discs, ligaments, tendons, muscle and other connective tissues (NICE, 2009).

Low back pain: Specific - Specific LBP would be classified as being caused by malignancy, inflammatory disorders, infections or fractures (NICE, 2009).

Manual therapy – is the movement of joints and /or soft tissue by the therapist hands. Manual therapy includes techniques including spinal manipulation, spinal mobilisation and massage (NICE 2009).

- *Spinal manipulation = 'a low amplitude high velocity movement at the limit of joint range taking the joint beyond the passive range of movement' (NICE, 2009).*
- *Spinal mobilisation = 'joint movement within the normal range of motion' (NICE, 2009).*
- *Massage = 'manual manipulation / mobilisation of the soft tissues' (NICE, 2009).*

National Institute of Health and Clinical Excellence (NICE) – An organisation in England and Wales, which provides national guidance and advice to improve health and social care (NICE, 2018).

Neuropathic pain – damage or disease affecting the somatosensory nervous system (Murnion, 2018).

Nociceptive pain – pain attributed to the stimulation of peripheral receptors (nociceptors) of the primary afferent neurons due to noxious stimuli (chemical, mechanical, thermal) of non-neural tissues (Smart et al. 2010).

World Health Organisation (WHO) – A specialised agency of the United Nations concerned with international public health (WHO, 2018).

1. Introduction

Low Back Pain (LBP) presents a major challenge worldwide; it affects individuals from every aspect of society and its economic and personal impact continues to grow, rising by more than 50% since 1990 (Clark and Horton, 2018; Cholangitis, 2018).

Lifetime prevalence in industrialised countries was reported at between 60 - 80% (World Health Organisation, 2013; Maniadakis and Gray, 2000) and global lifetime prevalence was reported at 40% (Hoy et al. 2012). The National Institute of Health and Clinical Excellence (NICE) (2016) proposes that LBP impacts upon more individuals globally than any other medical condition, and it is now reported as being the primary cause of disability globally (Hartvigsen et al. 2018; Vos et al. 2016).

LBP is the principal cause of activity limitation and work absenteeism throughout much of the world, impacting greatly upon global disability and the global economy (Hartvigsen et al. 2018; Thelin, Holmberg and Thelin, 2008; Lidgren, 2003).

Although in recent years the availability of medical interventions for LBP has increased significantly (Deyo et al. 2014) and there are a multitude of healthcare interventions and resource dedicated to LBP, there appears not to have been a corresponding reduction in the negative impact of LBP upon individuals and society (Foster et al. 2018; Deyo et al. 2014).

LBP is a complex disorder with multiple aspects; it is a heterogeneous condition encompassing a broad range of symptoms and pain presentations e.g. nociceptive, neuropathic or central pain (Nijs et al. 2015). Patients present to clinicians with a variety of symptoms, which are often not consistent, and currently there is great variation in clinicians' approaches to LBP's management (Duthey, 2013). Although national and international guidelines exist, to date, no consistent approach in the management of LBP has been accepted or adopted by clinicians locally or internationally (Duthey,

2013). O'Sullivan and Lin (2014) argue that if the management of LBP was more closely aligned with the evidence, the burden of LBP could be decreased.

Existing evidence and clinical guidelines support the use of diagnostic techniques, self-care, manual therapy, exercise and possibly acupuncture, to address LBP (Qaseem et al. 2017; Wong et al. 2017; NICE 2016; Nijs et al. 2015; Koes et al. 2010). Debate does however exist surrounding the role of acupuncture. Some international guidelines support the use of acupuncture, for example: America (Qaseem et al. 2017 and Chou et al. 2007), Scotland (SIGN, 2013) Canada (Cutforth et al. 2011), Belgium (Nielens et al. 2006), and previously England and Wales (NICE, 2009). Other international guidelines do not recommend acupuncture, for example: England and Wales (NICE, 2016), Europe (Airaksinen et al. 2006), France (Agence, 2000) and Italy (Negrini et al. 2006) and other guidelines overlook the intervention: the Netherlands (KNGF, 2013; CBO, 2003).

The UK's National Institute of Health and Clinical Excellence (NICE) LBP guidelines illustrate this variation. NICE supported the use of acupuncture in their 2009 publication (NICE, 2009) but withdrew support for acupuncture in their more recent guidelines (NICE, 2016) despite apparently no new contradictory evidence being produced; this policy change will be discussed further within chapter two. Due to the variable viewpoints surrounding the use of acupuncture for LBP, further research is required to establish the role of acupuncture and this thesis will investigate this further.

In this chapter, I consider the academic and, policy and practice context of the role of acupuncture and manual therapy for Low Back Pain (LBP). Acupuncture and Manual therapy were selected for this thesis in response to their recommendation in the 2009 NICE guidance; this will be discussed further in section 1.5. The specific parameters used to define LBP are reflected upon within this thesis and are outlined within section 1.1. The epidemiology and economic burden of low back pain upon the global society will be presented within section 1.2 and 1.3 respectively. The prognosis of LBP in section 1.4 and management of LBP will be presented in section 1.5.

An outline of the current evidence and clinical guidance for treatments available for LBP and how these have changed over time, with specific focus upon acupuncture and manual therapy will be given. Consideration to LBP being a complex condition, treated with complex intervention and the difficulties in assessing these will be discussed in section 1.6. This chapter will demonstrate a clear need for the research undertaken for this PhD, and the research questions it aims to answer will be presented in section 1.7. A thesis chapter plan will be outlined in section 1.8.

1.1 Definition: Low Back Pain

Defining the term 'Low Back Pain' (LBP) presents challenges. Establishing a definition and description for LBP is inherently complex and requires extensive explanation because as an applied medical diagnostic term, LBP is poorly classified. The World Health Organisation (WHO) states:

> *"Low back pain is neither a disease nor a diagnostic entity of any sort. The term refers to pain of a variable duration in the area of anatomy afflicted so often that it has become a paradigm of responses to external and internal stimuli"*

WHO (2003), page 671

While the term LBP is frequently used as a diagnostic term, it should perhaps be more accurately used as a diagnostic descriptor, a global term to describe a heterogeneous group of symptoms. Clinically LBP represents a collection of symptoms, such as pain, stiffness or discomfort in the region of the lower back, caused by a variety of structures and therefore the term LBP does not provide any specificity as to an exact anatomical structural or causal diagnosis. Low Back Pain (LBP) is defined as:

> *"...pain, muscle tension, or stiffness, localised below the costal margin to the inferior gluteal folds, with or without referred or radicular leg pain (sciatica)."*

McIntosh and Hall (2011), page 2

This definition provides no precision as to the actual anatomical structure or injury causing the condition; it only describes generalised sensations in a generalised region.

The NICE guidelines define LBP as:

> *"...the area bounded by the bottom of the rib cage and the buttock creases. Some people with non-specific low back pain may also feel pain in their upper legs, but the low back pain usually predominates. Several structures, including the joints, discs and connective tissues, may contribute to symptoms."*

NICE (2009), page 10

Consequently, there is a significant lack of clarity for generalised pain in the lumbosacral region and little consensus in defining LBP (NICE, 2016; Parthan, Evans and Le, 2006). This lack of clarity in the definition and thus diagnosis of LBP, may contribute to the variety that exists in the management, treatment and research of the condition (Duthey, 2013; Forward, Wallace and Hughes, 2008).

1.1.1 *Specific and Non-Specific Low Back Pain*

For research and clinical purposes, it is important to differentiate between two categories of LBP: Specific LBP and Non-Specific LBP, as they constitute quite considerably different conditions with varying care pathways as detailed below.

Initial diagnostic activity in relation to LBP is conducted by means of a thorough medical assessment. Clinicians utilise clinical red flag screening tools to exclude or identify any serious pathology. Identification of a serious or specific pathology would indicate the need for onward specialist referral and further investigations, whereas the exclusion of a specific pathology

would indicate a conservative treatment pathway for non-specific LBP (NICE, 2016; O'Sullivan and Lin, 2014; Duthey, 2013).

Specific LBP:

LBP is considered specific when a specific pathology is identified as a cause of the LBP. Specific LBP constitutes 5 - 15% of LBP cases within primary care in the UK (Hollingworth et al. 2002). LBP caused by serious pathology is very rare, less than 1% of LBP sufferers in primary care will receive a diagnosis of cancer for example (Henschke et al. 2013).

Serious LBP with a definitive diagnosis is referred to as 'Specific' low back pain, due to its specific diagnosis. It has been identified by NICE as:

> *"Specific causes of low back pain = malignancy, infection, fracture, ankylosing spondylitis and other inflammatory disorders"*

(NICE, 2009, page 10)

The diagnosis of 'specific LBP' encompasses a collection of conditions, which are generally more severe in their nature as identified above; they require complex invasive investigations and interventions by specialist physicians. The specific diagnosis allows for a distinct care pathway to be adopted within primary care for the identified condition.

The complex nature and requirement of intensive investigations, treatments and care pathways for specific LBP by their nature determine them to be beyond the remit of this study, they are often life-threatening conditions.

Non-specific LBP:

LBP is identified as non-specific because a specific cause cannot be conclusively established (Riksman, Williamson and Walker, 2011), it is suggested that it may occur due to a variety of biologic and behavioural influences (Deyo et al. 2014). Non-specific LBP accounts for the majority of LBP, with 85 – 95% of LBP being diagnosed as non-specific, not caused by a specific or identifiable pathology (Hollingworth et al. 2002).

While a variety of definitions exist for non-specific LBP, NICE's (2009) classification manages to encompass the key components of non-specific LBP, including the varied nature of structures involved and the exclusion of specific causes, they provide a descriptive narrative:

> *"Non-specific LBP is tension, soreness and/or stiffness in the lower back region for which it isn't possible to identify a specific cause of pain. Several structures in the back, including joints, discs, and connective tissues, may contribute to symptoms. The diagnosis of non-specific low back pain is dependent on the clinician being satisfied that there is not a specific cause for their patient's low back pain"*

(NICE, 2009, page 9)

Sensory nerves innovate several spinal structures within the low back, the sensory nerves allow the individual spinal structures to be pain generating and thus produce pain in the lower back. Currently despite the advancement in imaging technology and medicine, it is not possible to identify the specific causes of sensory nerve stimulated pain (NICE, 2016) and thus it is considered as non-specific.

Significant heterogeneity occurs amongst non-specific LBP, however duration-based categorisation has been used conventionally and continues to be used often in attempt to classify it further; three categories have traditionally indicated divergent treatment pathways for LBP (NICE, 2009):

- *Acute LBP – is classified as LBP incidence within the first six weeks*
- *Sub-acute LBP – is classified as LBP for between six to twelve weeks*
- *Chronic LBP – is classified as LBP lasting longer than six to twelve weeks or recurrent back pain that continues to affect an individual constantly or intermittently over a period of time*

(Duthey, 2013)

Though international medical pathways and extensive research studies have adopted these categories, the most recent UK NICE LBP guidelines (NICE, 2016) recommended adopting a new approach. They consider LBP as a continuum and recommend attempting to identify risk of recurrence and chronicity to categorise LBP, as opposed to duration-based time frames. In a break from convention they advocate identifying risk factors to be a more appropriate indicator of the risk of chronicity, and thus may guide care pathways and treatment requirement for LBP more appropriately, rather than categorisation of LBP by the duration of symptoms (NICE, 2016). This approach has not currently been adopted internationally.

In addition to the lack of diagnostic clarity, multiple terms are currently used interchangeably as descriptors for LBP without a specific cause; the clinical community have currently not converged upon one term (figure 1.1).

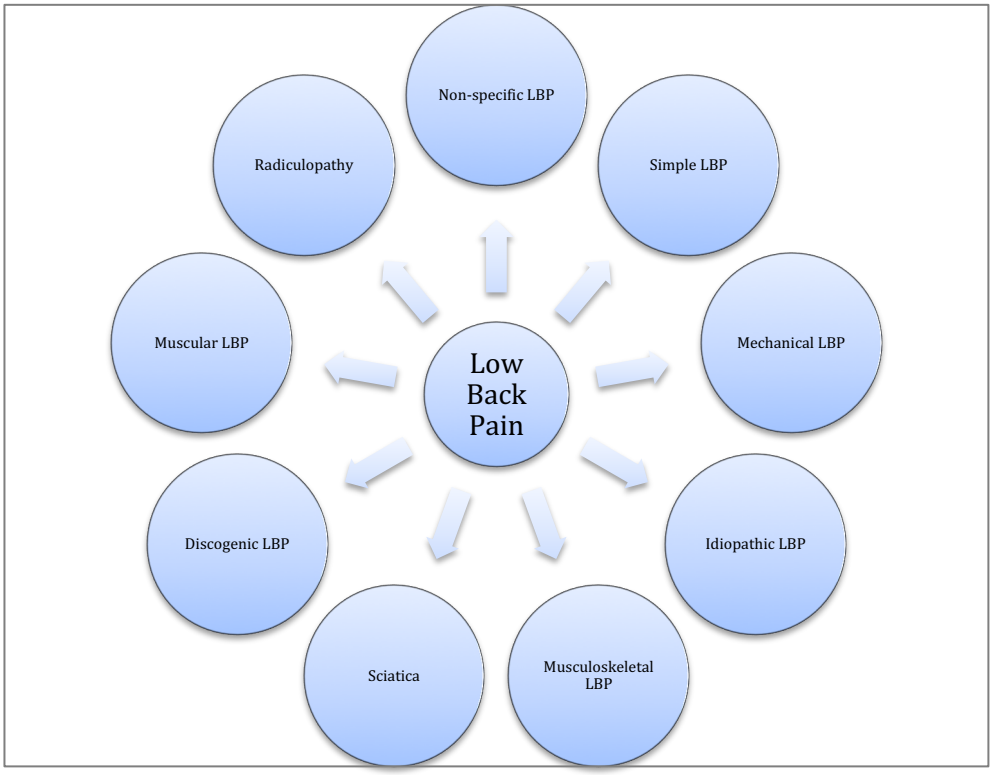

Figure 1.1: Diagnostic terms for LBP (adapted from NICE, 2016; Duthey, 2013; Forward et al. 2008)

The focus throughout this thesis will be to explore Non-specific LBP of at least a six-week duration (beyond the acute phase). For consistency and

clarity, where LBP is used within the thesis, it will be referring to sub-acute and chronic 'Non-specific LBP'.

1.2 Epidemiology of Non-Specific Low Back Pain

1.2.1 Aetiology and risk factors of Non-Specific LBP

Non-specific LBP can be described as idiopathic in nature with an individual being unable to identify a specific cause or incident resulting in their LBP. Identifying a specific anatomical pain source can thus prove problematic, in turn posing difficulty in diagnosis and subsequent treatment (Deyo et al. 2014; Riksman et al. 2011).

As LBP is considered as a pain in the region of the lumbosacral region of the spinal column, it can occur for a variety of reasons and from a variety of structures. The soft tissues involved in LBP include the muscles, vertebral discs, spinal and peripheral nerves, joint capsules, soft connective tissue, ligaments and blood vessels (NICE, 2016). Many of these structures are often involved with non-specific LBP with some or all being involved or affecting pain simultaneously within the course of an LBP episode. LBP is described frequently in medical practice as a result of overuse of a muscle, a strain of the soft tissues or a repetitive injury suffered over time (Parthan et al. 2006).

When a tissue structure becomes a trigger for the LBP through a sprain, strain, pull or stretch it becomes inflamed producing chemicals such as cytokines, these chemicals initiate the physiological process of swelling within the local tissues and its surrounding structures. Swelling, though required for the healing process can inadvertently limit blood flow and the delivery of key nutrients for healing. The swelling can also limit the flow and removal of inflammatory products, thus causing further irritation to the region. If inflammation is not reduced, a cycle of pain and inflammation can occur, leading to chronicity of pain (Duthey, 2013; Forward et al. 2008).

Though it is generally considered that the establishment of a specific cause for non-specific LBP is not possible, Deyo and Weinstein (2001) proposed that the anatomical origin of non-specific LBP could be sub-divided further into the following (figure 1.2):

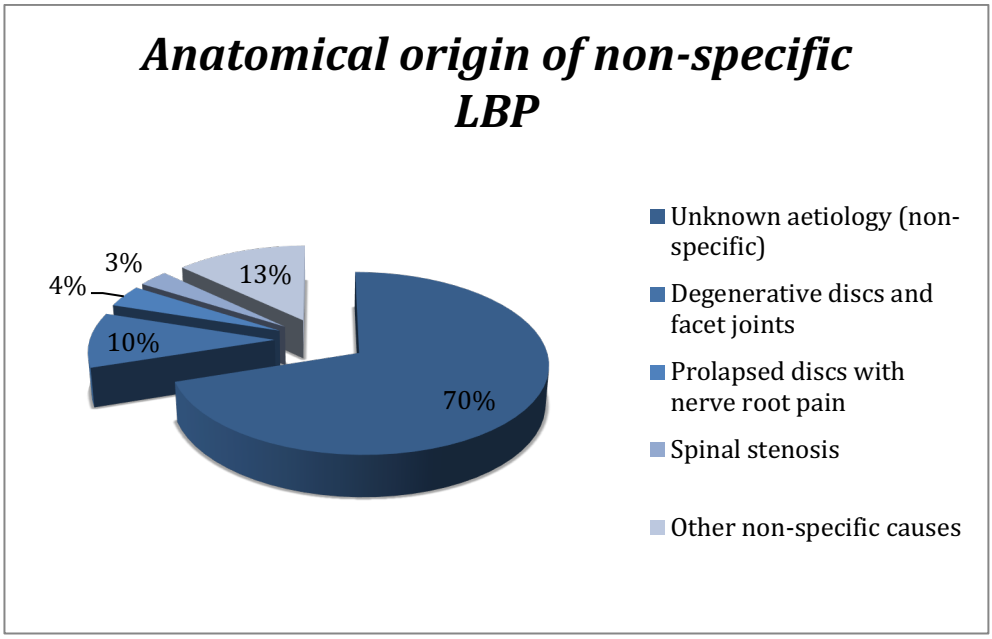

Figure 1.2: Anatomical origin of non-specific LBP (Deyo and Weinsein, 2001)

This approach however, was unable to establish a cause for 70%, which was categorised as 'unknown aetiology' for LBP (figure 1.2). Additionally, MRI and CT scans have determined facet joint disease, disc prolapses, and disc degeneration can often be a normal part of the aging process and non-symptomatic (Tonosu et al. 2017; Brinjikji et al. 2015; Lateef and Patel, 2009). The anatomical origin presented therefore, may not be the root cause of a person's LBP, thus limited information is provided to guide diagnostic causation of pain or in the guidance of following a specific treatment pathway.

Various risk factors have been associated with LBP; these include incorrect and sustained postures, manual work, trauma (i.e. car accident), sporting injuries, repetitive strains and socio-demographic factors including height, weight, lifestyle, age, smoking, physical fitness, and psychological influences (NICE, 2009).

Globally it was proposed that 37% of LBP could be attributed to occupational incidences. Various occupations expose individuals to sustained standing or sitting positions, (including occupational drivers), and activities involving vibrations or manual work with excess lifting, and are commonly linked with an increased risk of LBP (Duthey, 2013).

Inequalities within society have also been reported to impact LBP, with disadvantaged groups having a higher reported incidence of LBP, but receiving a lower onward referral rate for interventions or investigations than non-disadvantaged groups (DoH, 2006).

1.2.2 Global Burden of Disease

The Global Burden of Disease (GBD) study in 2017 included 345 diseases and injuries, including musculoskeletal disorders, from 195 countries. The global study provided estimates of the impact and burden of each condition. Of the 345 conditions studied, LBP was positioned globally as the leading cause for years lived with disability (YLD) and listed as second for overall burden of disease in terms of disability-adjusted life years (DALYs) in high socio-demographic countries (WHO, 2018). This is an increase in the burden of LBP from the 2010 GBD study when LBP was sixth in terms of DALYs.

Hoy et al. (2014) ranked the YDLs and DALYs for LBP globally and across individual regions for the 2010 study, demonstrating the burden of LBP internationally, and identifying the regional variation (table 1.1). LBP impacts greatly upon society internationally and it is suggested that when considering the reported number of cases and the burden of disease in terms of DALYs, LBP has a more profound impact upon society than HIV, road injuries, tuberculosis, lung cancer, and chronic obstructive pulmonary disease (Hoy et al. 2014; Lozano et al. 2012).

Region	YDL ranking	DALY ranking	Region	YDL ranking	DALY ranking
Globally	1	6	**Central Latin America**	2	7
Central Asia	2	7	**Southern Latin America**	1	2
East Asia	1	5	**Tropical Latin America**	1	3
Southeast Asia	1	2	**North Africa / Middle east**	1	2
Australasia	1	1	**North America High income**	1	3
Caribbean	4	13	**Oceania**	2	14
Central Europe	1	3	**Central sub-Saharan Africa**	3	23
Eastern Europe	1	3	**Eastern sub-Saharan Africa**	3	17
Western Europe	1	1	**Southern sub-Saharan Africa**	4	15
Andean Latin America	2	5	**Western sub-Saharan Africa**	2	13

Table 1.1: Regional LBP YDL and DALY ranking out of 291 conditions in 2010 (adapted from Hoy et al. 2014)

In a 2005 study of medical consultations (not solely emergency visits) in the United States of America (USA), LBP accounted for approximately 14 million consultations with physicians across the country (Snider et al. 2008). LBP was also estimated to account for 2.5% - 3% of emergency department visits within the United States, which equates to 2.63 million consultations (Friedman et al. 2010).

Within England there has been a growing trend of hospital admissions and surgical procedures for LBP, over a 15-year period, with the greatest increases seen in the older population (Sivasubramaniam et al. 2015) as detailed in the figure 1.3 and 1.4. The data was collected retrospectively using

Hospital Episode Statistics of lumbar spine disease cases in England between 1999 and 2013 (Sivasubramaniam et al. 2015).

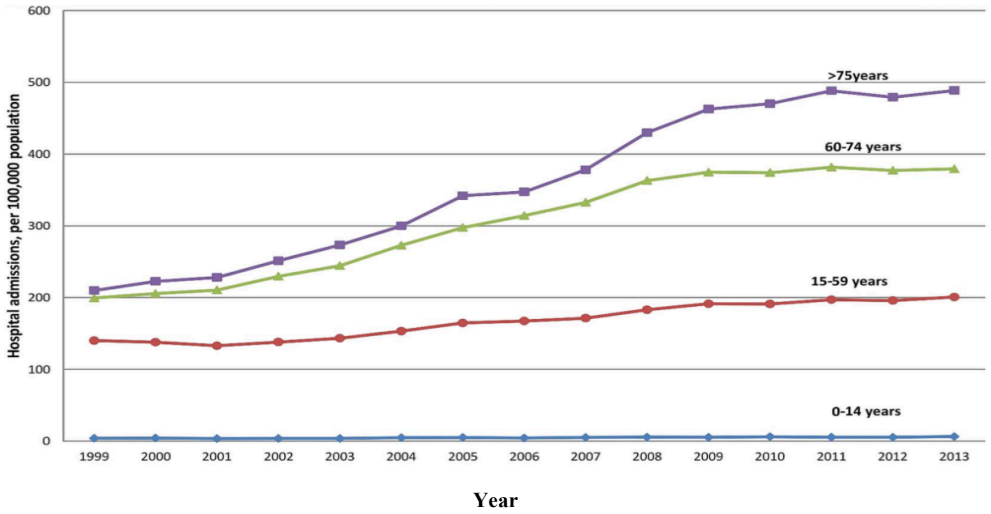

Figure 1.3: Hospital admission for LBP by age group in England (adapted from Sivasubramaniam et al. 2015)

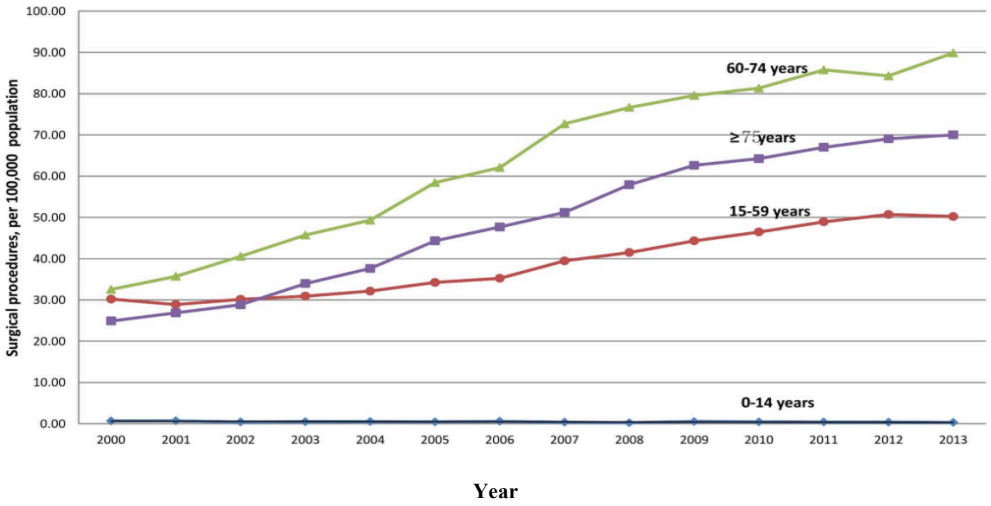

Figure 1.4: Number of surgical procedures for degenerative lumbar spine disease by age group in England (adapted from Sivasubramaniam et al. 2015)

Hospital admissions were reported to have risen from 127 to 216 per 100,000, with surgical procedures rising from 25 to 49 per 100,000, reflecting a significant cost and burden to the NHS and society within England (Sivasubramaniam et al. 2015).

1.2.3 Global Incidence of Low Back Pain

Incidence of LBP, the measurement of newly diagnosed cases within a specific time period, is frequently measured over a one-year period and relies greatly upon the accuracy of reporting. Measuring the rate of new cases treated may in some contexts be a more practical measure of incidence, as this figure is more often known and recorded, whereas the incidence of untreated or unreported cases is commonly unknown (Greenberg et al. 2005).

Comparing estimates of incidence can be problematic due to heterogeneity of information collected across studies, some measure first incidence of LBP and other studies include recurrent events. The expense of longitudinal studies required for incidence measurement means that fewer data studies are available for incidence, comparative to prevalence data detailed below within section 1.2.4 (Hoy et al. 2010).

International incidence:

Internationally the incidence of LBP is significant, there are variations from region to region and Western Europe has the highest incidence of LBP worldwide (Hoy et al. 2014).

In Germany a nationwide computer assisted telephone interview (CATI) of 8318 individuals indicated incidence at approximately one in four adults (25%) over the age of 30 years reports chronic LBP each year and LBP is one of the major causes of disability in Germany (Neuhauser, Ellert and Ziese, 2005).

In Denmark it is reported that although 60% to 90% of the population are expected to experience LBP during their life, the annual incidence is reported at 5%, considerably lower than other countries within Europe, however the data was collected in 1988 (Bekkering et al. 2003; Frymoyer, 1988).

No large longitudinal population studies existed within the USA, however using a National Electronic Injury Surveillance System of registered cases of LBP presenting to emergency departments, it was estimated that 2.06 million

cases of LBP were reported in a population of 1.48 billion, equating to an incidence of 1.4 per 1,000 person-years (Waterman, Belmont and Schoenfeld, 2012).

No international consensus on measuring the incidence of LBP appears to exist. It is difficult therefore to determine if the international differences in incidence of LBP are real, are likely linked to lifestyle and access to healthcare, or if the variation may in part be due to the variation in measurement and data collection approaches (Bohman et al. 2014).

UK incidence:

A population survey conducted by the Department of Health (DoH) in the United Kingdom (UK) in 1998, reported incidence of LBP at 40% within the preceding 12 months (DoH, 1999) this was calculated using data reported to the Health and Safety Executive (HSE) under the Reporting of Injuries, Diseases and Dangerous Occurrences (RIDDOR) scheme. While no further current UK data is available from the DoH specific to LBP incidence, within the UK over 200 different types of musculoskeletal disorders are registered and reported, and LBP was reported by the DoH to account for more than 40% of all musculoskeletal disorders as reported to the HSE within the UK (DoH, 2006).

More recent data are available from the UK Health and Safety Executive (HSE) (2017) on the working population; this data was collected using the Labour Workforce Survey (lFS) and RIDDOR reporting. A report on work related disorders in 2017 reported 194,000 (590 cases per 100,000 workers) cases of LBP across the labour workforce between 2016 and 2017, with an associated incidence rate of 150 per 100,000 (95% CI).

The National Institute for Health and Clinical Excellence (NICE) an organisation, which provides national guidance for healthcare improvement in England and Wales, has suggested that nearly everyone within the UK would be affected by LBP at some point within their lifetime (NICE, 2009. Cholangitis, 2018; Clark and Horton, 2018). A lifetime incidence of between 58 – 84% was estimated by Bernstein et al. (2017) and Parsons et al. (2011)

based on data from Dionne in 1999, collected via the 'House Survey for England' and the 'General Household Survey', a national survey conducted in England and Wales on behalf of the Department of Health (DOH). Milczarek (2009) reported approximately 62% of sufferers still live with their LBP after twelve months, contributing to the chronicity of the condition.

1.2.4 Global Prevalence of Low Back Pain

Prevalence is the measure of live cases of LBP at a specific time point and thus is more easily estimated than incidence. Period prevalence is the measure between two time points and point prevalence is measured at a specified time point (Greenberg et al. 2005). Cross sectional studies are used to measure prevalence. However analysing global prevalence over time is fraught with difficulty due to the complexities of obtaining accurate population estimates, significant methodological heterogeneity, variations in studies measurement of LBP and the recall periods required by participants across studies (Edwards et al. 2016; Hoy, et al. 2010; Konstantinou and Dunn, 2008).

International prevalence:

A systematic review (SR) of the global prevalence of LBP reported a total mean prevalence was 31.0% ±0.6%, with a one-year prevalence of 38.0% ± 19.4% and a lifetime prevalence of 39.9% ±24.3 (Hoy et al. 2012). The point prevalence of 11.9% ±2.0% was also reported with a one-month prevalence of 23.2% ±2.9% (following adjustments for inconsistencies in methodology), substantiating LBP as a significant consideration on a personal and a global level (Hoy et al. 2012).

Global lifetime prevalence data was collected by the Global Burden of Diseases Study and reported LBP at 40%, with Western Europe having the highest prevalence (Hoy et al. 2014). Lifetime prevalence increases significantly to 60 – 80% in western industrialised countries (WHO, 2013. Maniadakis and Gray, 2000). The range across countries is marked however; the point prevalence was estimated at 10% of the adult population in

Denmark, based the Danish National Health survey, vastly lower than the majority of Europe, (KJøller and Rasmussen, 1995; in Dufour et al. 2010). France reported an estimated prevalence of 50% with hospital fees accounting for a large portion of the cost of LBP based on a National survey and Hospital attendance data (Rossignol, 2009; Konstantinou and Dunn, 2008). Prevalence in Germany was unknown until a prevalence study was conducted in 2005 using a nation-wide computer-assisted telephone interview (CATI), it was reported one in four adults over the age of 30 years reported chronic LBP (Neuhauser, Ellert and Ziese, 2005).

There are various possible explanations for the international variations in LBP prevalence such as; varying reporting and recording of data, country specific interpretation of data, lifestyle and cultural differences, and variations in medical and self-care (Hoy et al. 2012; Hoy et al. 2010).

LBP affects a diverse range of populations across the globe, impacting upon the young through to the very old, although the prevalence and burden internationally increases during the sixth decade of life (Hoy et al. 2012; Breivik et al. 2006). Prevalence was found to be higher in men (mean 10.1%, 95% CI 9.4 – 10.7) than women (mean 8.7%, 95% CI 8.2 -9.3) with this pattern translating globally (Hoy et al. 2014).

UK prevalence:

The Department of Health (DoH) survey also reported prevalence within the UK, reporting a lifetime prevalence of 80% (DoH, 2006) no further current data is available from the DoH specific to LBP prevalence within the UK general population.

The HSE in their report on work related disorders in 2017, stated that of all conditions reported, LBP was the highest, with 590 cases per 100,000 workers reported from 2016 to 2017. The prevalence of LBP was reported to be greater for men than women in the working population (720 per 100,000, 580 per 100,000, from 2014 - 2017) and a total of 194,000 cases of LBP were reported within the UK workforce between 2016 and 2017 (HSE, 2017).

While the majority of data is derived from the working-age population it is suggested that LBP is most prevalent in the seventh decade of life in the England and Wales (NICE, 2016).

1.2.5 LBP epidemiological data

A dearth of epidemiological data on LBP has long existed, however in 2000 epidemiological research into the area of LBP was reported to be growing (Walker, 2000). In 2010, Hoy et al. stated that further research was required into the epidemiology of LBP, they also established that there was a need for epidemiological studies to adopt a more standardised approach to each other across the research spectra to aid data comparisons of LBP epidemiology internationally (Hoy et al. 2010).

These views were echoed again in 2015 and recommendations for epidemiological studies of LBP to follow the National Institutes of Health (NIH) pain consortium task force guidance were encouraged (Deyo et al. 2014), with minimum data sets advised. A lack of research was also identified within developing countries, with the majority of studies focused upon European society with Caucasian individuals, reducing the global generalizability of studies and current data sets (Meucci, Fassa and Faria, 2015).

The lack of standardised tools for diagnosing LBP, further complicates data collection; often research data of LBP relies upon self-reported episodes (Parthan et al. 2006). Significant heterogeneity in data collection and objective measures continues to be present across LBP studies globally causing international comparisons to be complex (Hoy et al. 2014).

There continues to be limited epidemiological research specific to LBP relative to the significant burden it places upon society. It is also suggested that though LBP is widespread and impacts greatly upon individuals and society, it continues to be relatively under-funded and under prioritised for treatment, investment and research comparative to other conditions of similar or lower prevalence (Buchbinder et al. 2018). This may possibly be explained

by LBP often being considered comparatively trivial, as it is not a life limiting condition (Hoy et al. 2014; Hoy et al. 2010).

1.3 Economic Burden of Low Back Pain

The economic burden of LBP is the estimated total associated cost to individuals and society; including any resource use which would not have occurred if the condition were not present (Dagenais, Caro and Haldeman, 2008). The costs of LBP can be divided into three categories; direct costs, indirect costs and intangible costs (Odole et al. 2011).

- *Direct costs: include the cost of medical appointments, investigations, medications, interventions and all associated medical costs for a condition.*
- *Indirect costs: incorporate productivity loss for industry, earnings lost to individuals and family members, cost of paid support, transportation for appointments and sustenance purchases, premature death may also be incorporated into this calculation for disease limiting indirect costs.*
- *Intangible costs: are often unquantifiable and difficult to measure, but can include the effect of disease on third party individuals, examples could include, if an individual were on long term sick with LBP, this can impact on the morale of the colleagues having to cover additional workload, or activities or important events missed by family members due to parental incapacity with LBP.*

(Odole et al. 2011)

International economic burden:

The burden of LBP upon international communities and economies is considerable (Costa et al. 2009). The WHO (2013) determined LBP to be the leading cause of activity limitation and absence from work throughout the world. As with other healthcare measures the measures of burden upon

economies between countries and within individual countries is varied, with limited consistency in measurement of data (Hoy et al. 2010).

In the USA in 2002 annual direct health care costs related to back pain care specifically were estimated at $17.7 billion, (Mychaskiw and Thomas, 2002). An alternative estimate calculated the economic cost to be $105.4 billion (Van der Roer et al. 2005). Both these estimates used data from the Medical Expenditure Panel Survey (MEPS) and the principle explanation for the difference in estimates, is the latter study reported all healthcare costs for individuals suffering LBP, as opposed to direct healthcare cost specifically attributable to LBP (Parthan et al. 2006). An alternative estimate of direct healthcare costs of LBP was $85 billion dollars nationally in 2005, which reflected a vast increase from the 2002 $17.7 billion estimate and a reported 65% increase in cost from 1997 (Martin et al. 2008). In 2010 it was suggested that LBP was the fifth most common reason to access health care in the USA with an annual spend of $30 – $50 billion on LBP treatment specifically (Waterman et al. 2010; Chou and Chekelle, 2010).

Lost revenue in the USA, due to indirect costs and lost working days was estimated at $22 billion in 2005, this estimate included medication costs, medical consultations and hospitalisation (Parthan et al. 2006), while an earlier estimate of the annual productivity loss was at $34 billion, the estimate included costs of health care utilisation, wage estimates and productivity and disability (Rizzo, Berger and Abbott, 1998). Additional to lost revenue approximately 2% of the work forces within the USA were financially compensated for their occupational back pain injuries annually (Della-Giustina, 2015). The total inclusive costs related to LBP were estimated to be in excess of $100 billion per year (Crow and Willis, 2009).

In Germany LBP was reported as one of the major causes of work absenteeism and disability in the country (Neuhauser et al. 2005). Another German study measured their economic costs on a per patient basis, with annual direct costs estimated at 7000 euros per person for LBP sufferers and work absenteeism accounting for 75% of total per patient cost of LBP in Germany (Juniper, Le and Mladsi, 2009).

In Sweden a study calculated that in addition to direct costs, the indirect costs were extensive for those with LBP. Direct costs were estimated at 3,100 euros, whereas indirect costs were estimated at 17,600 euros per LBP patient (Ekman et al. 2005).

Maniadakis and Gray (2000) stated over 17 years ago there was a lack of economic and incidence data on LBP, a condition whose impact was so extensive and costly. Though further data has become available there continues to be limited exploration on this subject and a lack of consistent measurement and reporting of economic data (Hoy et al. 2014).

UK economic burden:

LBP accounts for 11% of the total disability within the UK population, implicating LBP as the greatest singular source of incapacity of a non-fatal condition within the UK (Greenough, 2016; Sivasubramaniam et al. 2015).

Year	Estimated Cost	Area measured	Reference
1998	£1632 million	Direct healthcare costs to the NHS	Maniadakis and Gray, 2000
1998	£11 billion	Indirect and direct costs in the UK	Maniadakis and Gray, 2000
2001	£12.3 billion	Associated societal cost. (2.5 million individuals in the UK were reported to have back pain on every single day of the year, 22% of UK healthcare spend)	PCC, 2010
2002	£69 million	Retrospective cohort study in UK primary care, determined 4.6 million GP appointments (equivalent of 793 fulltime GP roles)	Belsey, 2002
2003	£481 million £197 million	Per year to NHS Private consultations	PCC, 2010
2007	£1.1billion	NHS	Critchley et al. 2007

Table 1.2: Estimated cost of low back pain in the UK

Economic cost and healthcare cost estimates vary greatly across studies due to their dependence upon cost-of-illness approximations, fundamental suppositions, and the differing study methodologies and measurement analyses used across the data sets (Phillips, 2006). Various annual estimates have been identified, detailed in table 1.2; however there appears to be a lack of agreement. The estimated cost discrepancies may in part be due to the varying data collection methods used.

Current reported NHS cost data for LBP within the UK is lacking, however economic analysis data is available from other organisations and government departments, detailed in table 1.3.

Year	Estimated Cost	Area measured	Reference
2017	3.2 million working days lost due to LBP, 16.5 working days were lost on average to each LBP case	Reported 194,000 cases of LBP existed within the labour workforce in 2016/17	The Health and Safety Executive (HSE)
2008	33% of LBP sufferers had a recurrence of symptoms that caused their repeated absence from work	An occupational survey in Northern Ireland	NIAO, 2008
2007	Five million working days were lost each year. Half a million people receive a long-term state incapacity benefit	Chartered Society of Physiotherapy (CSP) report	CSP, 2007
2003	£1.4 billion estimated the cost of benefits paid to individuals afflicted by LBP. £3.8 billion additional lost productivity estimated	The Department for Work and Pensions (DWP)	PCC, 2010

Table 1.3: Estimated economic cost of low back pain in the UK

No further updates on LBP have been reported or are available from the Office of National Statistics (2017) or NHS digital (2017).

An often-additional cost and impact upon society are what are described as "hidden costs", an example of this is the nature and quality of care provided in general practice being variable across the UK. Some patients are referred

to physiotherapists, orthopaedic consultants or to a pain management team, while others are given advice and prescribed medication with no onward referral (Fritz, 2012). The implications for this variability to subsequent healthcare costs for delayed onward referral and thus chronicity of conditions is unknown, however it is suggested that the costs could be extensive (Care Quality Commission (CQC), 2017).Despite the extent and impact of LBP within the UK, there appears to be inconsistent measurement of LBP and its economic burden upon society. Consistent data records would provide more accurate and current measures of the economic burden of LBP; however, burden of disease is often not a principle driver for decision making, data on the cost of improving quality and length of life is more usually focused upon (Kennelly, 2017).

Interestingly though of all the economic analysis studies undertaken of all conditions, it was reported that LBP was found to be one of the costliest conditions to the UK economy. A finding also reflected internationally (Fritz, 2012).

1.4 Prognosis of Low Back Pain

Individuals with their first incidence of LBP have a relatively good prognosis, with 60 – 70% recovering within six weeks and 80 – 90% improving within twelve weeks. The 10 – 20% who do not recover by twelve weeks (chronic LBP) have an uncertain outcome, with treatment and prognosis amongst this group being variable (Forward et al. 2008; Anderson, 1999; Frank et al. 1996).

Chronic LBP sufferers have a poor prognosis of full recovery, with 62% having LBP for beyond twelve months (Milczarek, 2009). Recurrent injuries occur for these individuals and 33% experiencing a recurrence requiring their absence from work (Milczarek, 2009; NIAO, 2008).

1.5 Management of Low Back Pain

Accessing the appropriate treatment at a suitable time can be key to improving the prognosis of an individual's LBP, and extensive literature and advice is available to encourage self-care and treatment, for LBP (CSP, 2017; Fritz, 2012). However, accessing the most appropriate advice specific to an individual can be inherently difficult, due to limited access, limited knowledge and the variety of information available (CQC, 2017). As mentioned previously it is noted that those from disadvantaged groups have been reported to be less likely to be offered treatments or diagnostic interventions and thus suffer a poorer prognosis (Hartvigsen et al. 2018; Sharim et al. 2017; DoH, 2006).

If individuals attempt to access treatment independently of the NHS, there are various professionals who offer treatment for LBP e.g. doctors, physiotherapists, chiropractors, osteopaths, acupuncturists, massage therapists, healing therapists (NICE, 2016) and various treatments are available e.g. exercise, education, physiotherapy, manual therapy, ultrasound, acupuncture, CBT, magnet therapy, hypnosis (Parthan et al. 2006). These diverse disciplines and treatments offer techniques with varying underlying principles, evidence and likely effectiveness; however, the public may find it difficult to discern which is the appropriate treatment to aid their condition. On-line access to information and social media provides greater information resource but this can also be misleading in guiding treatment and prognosis, as presented information may inflate claims of effectiveness or be untrue (Advertising Standards Authority (ASA), 2010).

Several treatment interventions are researched, available and used globally for LBP (see figure 1.5).

Figure 1.5: Interventions for LBP (adapted from Parthan et al. 2006)

Comprehending the vast expanse and quality of research available for all LBP interventions is impossible for professionals, many of whom rely upon systematic reviews and clinical guidelines to summarise the key information (Vale et al. 2015).

In England and Wales, the NICE guidelines are one source of information often used to guide clinical staff in relation to best available evidence and practice. NICE guidelines for LBP were produced in 2009 and updated in 2016 (NG88 2009 & NG59 2016). The change in recommendations between

these two guidelines and the removal of acupuncture from the 2016 guideline recommendations is discussed further in Chapter two.

1.6 Combining a Complex Condition with Complex interventions

LBP has long been considered a complex condition with multiple facets. As such, some argue that it should be treated aggressively using a multidisciplinary approach, with treatment aiming to address all aspects of the condition including the physical symptoms, psychological aspects and functional restoration (Salerno, Browning and Jackson, 2002).

Many of the individual interventions used for LBP are also considered complex in their nature and teasing out the therapeutic effectiveness of specific single interventions remains challenging. Defining the actual components (which are active, and which are inert) of a complex intervention is possibly the most challenging part of evaluating a complex intervention (Campbell et al. 2000).

Complex interventions are used widely within healthcare and are considered as interventions with several interacting components. Acupuncture and manual therapy are two such interventions, which will be considered within this thesis due to their recommendation in the NICE LBP guidelines in 2009 as presented in chapter two, section 2.6. Evaluating the complexities of these interventions is challenging for researchers with many RCT methodologies lacking the sensitivity to comprehend the causal relationships and their intricate associations (MRC, 2008).

The Medical Research Council (MRC) established a framework and guidance for developing RCTs for complex interventions in 2000 and updated their guidance in 2006 (MRC, 2008), the MRC are currently updating the guidance and this is due to be published in 2019 (Skivington et al. 2018). The adoption of this framework is widely considered to be good practice when researching the development and effectiveness of complex interventions and the framework was current at the time of conducting the trial for this thesis. The

MRC guidance highlighted the concept of 'Process Evaluation' of complex interventions, partly due to the effect sizes within RCT's not providing policy makers with substantial information for replication in clinical practice (Moore et al. 2015). Recognising the importance of process evaluation was established within the updated MRC guidelines, as it:

> *"can be used to assess fidelity and quality of implementation, clarify causal mechanisms and identify contextual factors associated with variation in outcomes."*

(Craig et al. 2008, page 3)

Prior to the development of the MRC framework, researchers generally focused upon a uni-faceted approach, often comparing single interventions to usual care, various forms of placebo or considering a single intervention in head to head comparisons for LBP.

Understanding complex interventions is an on-going area of research and development. The MRC recognises the need to understand how complex interventions work and are currently funding investigations of causal models of complex interventions in an attempt to determine how the mechanisms of action of interventions occur and how the variables of a mechanism influence one another (MRC, 2018).

In clinical practice, healthcare professionals rarely deliver single interventions but rather a package of interventions. The combination approach adopted by many is commonly considered more useful and may incorporate for example; exercise, advice, manipulation, massage and acupuncture, all within a single treatment session (Salerno et al. 2002). Establishing which components are therapeutic and which are inert continues to be open to debate (Gleiss, 2017). Adopting a pragmatic inclusive approach to the research process for LBP and combining interventions within research studies may deliver results more closely mapped in clinical practice. Thus far current research design may be failing to reflect current clinical practice (Salerno et al. 2002) and has not provided definitive statement on treatments

to enable the development of a comprehensive, effective pathway of care for LBP.

A complex condition like LBP may be suited to a multi-faceted package, combining complex interventions; this may be a more useful approach and may prove to be a more successful resolution for the treatment of LBP (Salerno et al. 2002). Therefore this thesis and the associated pilot RCT will investigate the comparison and combination of acupuncture and manual therapy treatments for LBP; this will be outlined in chapter four.

1.7 Conclusion

This chapter has provided an introduction and described the condition of LBP, highlighting its burden upon society. The impact of LBP both on the population and the economy internationally is substantial (Clark and Horton, 2018) and necessitates the need for further investigation.

With the aetiology of LBP frequently remaining unclear, a variety of treatment options are currently adopted by clinicians (Parthan et al. 2006), gaining some clarity regarding the most effective treatment pathways for LBP through further research would be clinically useful.

This chapter has started to build the case for continued research into the area of LBP. It has highlighted the need to consider research design for complex interventions, which may also better reflect clinical practice. Considering an approach to incorporate both acupuncture and manual therapy individually and combined may provide guidance into the most effective and clinically useful interventions for LBP and this will be explored in chapter two and three.

1.8 Thesis Research Questions:

- *Are acupuncture and manual therapy and their combination approporiate and viable treatments of LBP?*

- *What are the international LBP recommendations for acupuncture and manual therapy?*
- *Which RCT design could best evaluate acupuncture and manual therapy in the treatment of LBP?*
- *Is high quality research available comparing and combining acupuncture and manual therapy for LBP?*

1.9 Structure of the thesis

Chapter one: I discussed LBP, its epidemiology, economic burden, causes and prognosis. I described some of the current treatments available for LBP. I outlined the complexities of assessing complex interventions for a complex condition and concluded with key research questions to be considered throughout this thesis.

Chapter two: In this chapter I present a review of systematic reviews of acupuncture and manual therapy for LBP and investigate if any available evidence compared or combined acupuncture with manual therapy. I also review the systematic reviews of international clinical practice LBP guidelines, and the source international clinical practice guidelines, considering them in relation to the UK NICE guidelines and their recommendations specific to acupuncture and manual therapy. The chapter critically examines the NICE LBP guidelines from 2009 and 2016, and I report upon the changes over time, and their evidence and recommendations of manual therapy and acupuncture. I discuss the removal of acupuncture from the recent NICE guidelines (2016) and the uncertainty around its provision. In this chapter I start to consider if acupuncture and manual therapy are appropriate interventions for the treatment of LBP.

Chapter three: I consider the strengths and weaknesses of RCT designs and the most appropriate design for a trial of acupuncture and manual therapy for LBP. I provide the rationale for the use of a cohort design with a nested factorial RCT, exploring the impact of retention and recruitment to trials. Discussion and justification for using an active control group rather than a

placebo or sham arm within the trial is also given. In this chapter I establish the basis for deciding which RCT design could best evaluate acupuncture and manual therapy for the treatment of LBP.

Chapter four: In this chapter I justify the need for a pilot study, the professional group to take the lead on a study of LBP and which LBP outcome measures to utilise. I detail the aims, methodology and scientific procedures of the cohort investigation and embedded pilot RCT, which forms a major part of this thesis. I present the recruitment rates, study documentation, justification for the choice of study population and inclusion and exclusion criteria. The randomisation process and trial interventions are detailed, and the monitoring and methods of analysis are also presented.

Chapter five: I report the results of the pilot study and the use of the cohort study for recruitment to the nested factorial RCT. Providing a descriptive review of the information and participant data gained and the outcomes of the trial. I present an exploratory analysis, though it was not planned for the pilot to be powered to demonstrate effectiveness within the RCT.

Chapter six: In this chapter I discuss the results of the pilot study. I consider recruitment for the trial, acceptance and retention rates and any issues arising from the RCT. I discuss the results of the primary and secondary outcomes, the exploratory regression analysis and the comparison analysis of the two outcome measures, the results of the ancillary data and the sample size calculation. I discuss the design of the RCT for evaluating a population with LBP, the new findings from this pilot and I conclude in a discussion of the strengths and weaknesses of the RCT and what impact this information will add to future investigations in this area of research.

Chapter seven: I present a systematic review of acupuncture and manual therapy for LBP, which includes a meta-analysis including the data from the pilot RCT conducted for this thesis. I present a GRADE quality assessment and recommendations, and a discussion of the available research and its quality. The strengths and limitations of the review and the need for further evidence on acupuncture and manual therapy are discussed.

Chapter eight: Constitutes a summary of the thesis, I discuss key findings and identify limitations of the thesis and the pilot RCT. I reflect upon proposed future research identified from this study, highlight what has been contributed to the research knowledge base and consider what a future study may constitute. The chapter ends with conclusions from the thesis.

2 A literature review of acupuncture and manual therapy and a review of Clinical Guidelines for LBP

2.1 Introduction

In chapter one, I introduced the subject of LBP and laid the foundation for the study of this thesis. In this chapter I present a review of systematic reviews (SRs) of acupuncture and manual therapy for LBP and investigate if any available evidence compared or combined acupuncture with manual therapy, to inform a potential pilot trial for this thesis. I also present the current clinical LBP guideline recommendations internationally and I contextualise the international guidelines with the UK NICE LBP guidelines. I critically examine the NICE guidelines for LBP from 2009 and the updated version in 2016, and I discuss the disparities between the NICE guidelines, explore changes over time and reflect upon the most recent recommendations for the management and treatment of LBP. Specific focus of the evidence provided for acupuncture and manual therapy is given.

2.2 Literature review

2.2.1 Aims and methods of literature review

Aim of the literature review:

- *To review systematic reviews of acupuncture and manual therapy for LBP and to investigate if any available evidence compared or combined acupuncture with manual therapy for LBP*

Prior to conducting a pilot trial for this thesis, I conducted a literature review in March 2010 from the databases listed (Box 2.1). The plan of this literature review was to inform a potential planned pilot trial of the current evidence base. For the purpose of presenting this thesis, the search conducted in 2010

of systematic reviews, was repeated in 2018, to ensure all information provided in this thesis was updated, current and accurate.

Box 2.1: Sources searched for previous systematic reviews of acupuncture and manual therapy, and their comparison or their combination for the treatment of LBP:

- *The Cochrane Library*
- *The Cochrane Database of systematic Reviews*
- *The Database of Abstracts of Reviews of Effects (DARE)*
- *The Cochrane Database of Methodology Reviews*
- *Health Technology Assessment (HTA) Database*
- *NHS Economic Evaluation Database*
- *PROSPERO International prospective register of systematic reviews*
- *NIHR Health Technology Assessment Website* http://www.ncchta.org
- *Centre for Reviews & Dissemination Website* http://www.york.ac.uk/inst/crd
- *Pubmed (Medline)*
- *The internet, including Google scholar*

(Montori et al. 2005)

This search was conducted to establish the current available evidence. All SR's found, which considered acupuncture or manual therapy individually for LBP are detailed in sections 2.2.2.1 and 2.2.2.2 (tables 2.1 and 2.2).

Since the aim of my thesis was to investigate if acupuncture versus manual therapy and their combination could be considered as effective treatments for LBP, the databases were also searched for systematic reviews, protocols and publications of acupuncture versus manual therapy, and acupuncture combined with manual therapy for LBP (section 2.2.2.3).

I wanted to identify and explore if any existing systematic reviews, protocols or studies comparing or combining acupuncture with manual therapy were registered or currently planned, to avoid duplication and derive any lessons learned from the previously conducted systematic reviews. Any similar or associated reviews were reviewed to see if they were current, comprehensive (including all the available evidence) and to review any limitations.

2.2.2 Results of the literature review

My literature review revealed that although reviews existed for acupuncture and manual therapy used separately for LBP, no systematic reviews compared acupuncture with manual therapy or combined acupuncture with manual therapy for LBP (section 2.3.3). This indicates a potential need for further research in this area.

The SRs identified were checked to see if any included RCTs compared or combined acupuncture with manual therapy for LBP and if any existed, they were reviewed and reported upon.

2.2.2.1 Summary of systematic reviews of acupuncture for LBP

Table 2.1 summarises the SR's I found and further detail is provided within the narrative of this chapter.

Authors and Date	Title	Parameter dates	Number of Studies	Authors conclusions
Liu, Skinner, McDonough, Mabire, Baxter (2015)	Acupuncture for LBP: An overview of SR's	From database inception to February 2014	16 SRs	Acupuncture as an adjunct to conventional therapy provides short term pain relief and functional improvement, further research required
Zeng, Chung (2015)	Acupuncture for chronic nonspecific low back pain: An overview of systematic reviews	January 2003 to May 2014	17 SRs	Appears to be strong evidence for the effectiveness and moderate evidence for the cost effectiveness for the use of acupuncture in LBP
Lam, Galvin and Curry (2013)	Effectiveness of acupuncture for nonspecific chronic low back pain: a systematic review and meta-analysis	From database inception to May 2012	25 RCTs	Reported acupuncture was favourable for self-reported pain however exerted caution in interpreting the results
Lee, Choi, Lee, Lee, Shin, Lee, (2013)	Acupuncture for acute low back pain: a systematic review	From database inception to June 2011	11 RCTs	Determined the evidence was encouraging for favouring acupuncture over medication or sham but determined further evidence was required
Furlan, Yazdi, Tsertsvadze, Gross, Van Tulder, Santaguida, Gagnier, Ammendolia, Dryden, Doucette, Skidmore, Daniel, Ostermann, Tsouros, (2012)	A Systematic Review and Meta-Analysis of Efficacy, Cost-Effectiveness, and Safety of Selected Complementary and Alternative Medicine for Neck and Low-Back Pain	From database inception to February 2010	147 RCTs	Demonstrated short to medium effects of the CAM interventions compared to usual care, inconclusive results when compared to other interventions and negative results when compared to sham
Standaert, CJ. Friedly, J. Erwin, MW. Lee, MJ. Rechtine, G. Henrikson, NB. Norvell, DC. (2011)	Comparative effectiveness of exercise, acupuncture, and spinal manipulation for low back pain	From database inception to December 2010	1 RCT	Concluded insufficient evidence in this area
Furlan, A. van Tulder, MW. Cherkin, D. Tsukayama, H. Lao, L. Koes, BW. Berman, BM. (2005)	Acupuncture and dry-needling for low back pain	1996 to February 2003	35 RCTs	No firm conclusions of the effectiveness of acupuncture, some evidence for acupuncture being a useful adjunct to other treatments for LBP
Ernst, White (1998)	"Acupuncture for Back Pain: A Meta-Analysis of Randomized Controlled Trials."	From 1969 to 1996	12 RCTs	Concluded acupuncture superior to various control interventions, insufficient evidence of superiority to placebo

Table 2.1: Summary details of acupuncture SR's included in literature search

Two reviews (SR) were published in 2015. Liu et al. (2015) conducted a computer-aided systematic literature search of databases, and reviewed SRs

of acupuncture for LBP to February 2014. They included 16 SRs, and extracted the methodological qualities of the original RCT studies with judgement from the SR authors. The external validity of the SRs was assessed using the revised STRICTA reporting guidelines (MacPherson et al. 2010) and quality assessment was conducted using an AMSTAR checklist. They reported overall the methodological quality was low and external validity was weak of the included SR and the primary studies. This review presented comprehensive methods and report of the results. They concluded acupuncture as an adjunct to conventional therapy provides short term pain relief and functional improvement, but concluded further knowledge was required, and additional effort required to improve the quality of both RCTs and SRs in this area (liu et al. 2015) their conclusions appeared appropriate to the literature presented.

Zeng and Chung (2015) searched data from January 2003 to May 2014, for reviews published in English or Chinese in the Cochrane library, Allied and Complementary Medicine Database, Scopus and the Chinese Academic Journal (CAJ). They included 17 SRs and assessed the methodological quality of the reviews using the Overview Quality Assessment Questionnaire (OQAQ) (the 17 SR were rated as 7 x 9/9, 3 x 7/9, 3 x 6/9, 1 x 5/9, 1 x 4/9 and 2 x 2/9). They reported all the SRs held positive findings for acupuncture in LBP, and concluded "there appears to be strong evidence for the effectiveness and moderate evidence for the cost effectiveness of the use of acupuncture". However the review presented limited methodological information, and while they presented the evidence of varying quality of the studies, they did not report how the quality of the studies impacted upon their conclusions. Thus their positive conclusions of strong evidence for acupuncture for LBP appeared to be overstated, as only small to moderate effect sizes were reported, at least two SRs had presented inconclusive findings and one of the most recent SRs (Hutchinson et al. 2013) reported no difference between acupuncture and sham acupuncture in their review.

Lam, Galvin and Curry (2013) reviewed research to May 2012 and examined 25 individual studies. They used the Cochrane risk of bias tool and identified

low methodological quality in many of the included studies and substantial heterogeneity across the studies. Comprehensive methods, results and meta-analysis were presented. They reported acupuncture had clinically meaningful reductions of self-reported pain compared to sham (mean difference 16.75, 95% CI, -33.33 to -0.19), and improved function when compared to no treatment (standard mean difference = 0.94, 95% CI, -1.41 to -0.47), however they exerted caution in interpreting the results due to the limitations and the heterogeneity of the included studies. The conclusions of this SR appeared to be a fair representation of the RCTs included.

Lee et al. (2013) conducted a SR for acute LBP and identified 11 RCTs for consideration. They reported significant methodological limitations in the RCTs. The methods and results reported were reasonable but a meta-analysis was not conducted. They reported that the evidence was encouraging for favouring acupuncture over medication or sham but they recommended that further evidence was required. The author's findings were positive and broadly representative of the RCTs reviewed.

Furlan et al. (2012) conducted an extensive SR considering several areas of complementary and alternative medicine for low back and neck pain. This SR considered studies published until February 2010, including 147 studies and the analysis reported pain intensity and disability. Risk of bias was assessed using the Cochrane Collaboration Back Review Group (CCBG) criteria and their methods and results were comprehensively presented. The review included 33 RCTs of acupuncture and 23 manual therapy (manipulation and massage) RCTs but noted the quality of the studies was low. The results demonstrated short to medium effects of the CAM interventions compared to usual care, inconclusive results when compared to other interventions and negative results when compared to sham interventions. Within the SR, one study of acupuncture versus manipulation for LBP (Giles and Muller, 2003) and its pilot study (Giles and Muller, 1999) were reported and their data were pooled as a sub-analysis of this review, it concluded that manual therapy was superior to acupuncture. The SR was

comprehensive, including many types of intervention and the conclusions appeared to be a fair representation of the RCTs included.

Hutchinson et al. (2012) reviewed acupuncture research for low back pain from 1950 to 2011, they identified 82 RCTs but only seven were included in the SR based on their inclusion criteria, reporting 72 studies were excluded based on including pregnant ladies and the elderly in their RCTs. Comprehensive information on the methods and results were presented, though no meta-analysis was conducted. They indicated there was some evidence to suggest acupuncture was more effective than no treatment, but reported no conclusions could be drawn about its effectiveness over other modalities due to the conflicting evidence and poor quality of the studies. The author's findings appeared to be representative of the RCTs reviewed.

In 2011 a review attempted to investigate RCTs that compared the effectiveness of exercise, acupuncture, and spinal manipulation for low back pain. The review searched studies on MEDLINE and The Cochrane Library up to December 2010. They found one study comparing exercise to spinal manipulation; however, they were unable to find evidence that compared acupuncture with spinal manipulation or exercise and concluded there was insufficient evidence in this area (Standaert et al. 2011).

The systematic reviews listed below (also see table 2.1), were available in the March 2010 literature review and helped inform the planned pilot study for this thesis.

Furlan et al. (2005) investigated the treatment of acupuncture for LBP from 1996 to February 2003, they included 35 RCT with comparators of placebo, other interventions, no treatment or alternative acupuncture. The methods and reporting of the SR and meta-analysis was comprehensive and concluded no firm conclusions could be made regarding the effectiveness of acupuncture, but found some evidence for acupuncture being useful in adjunct to other treatments for LBP. Their conclusions appeared representative of the RCTs presented.

An earlier systematic review and meta-analysis by Ernst and White, 1998, of acupuncture for LBP searched databases from 1969 to 1996; 12 RCTs were included. The methods and reporting of the SR and meta-analysis was comprehensive, while they identified an assessment of quality criteria there was no report of the assessment results. It concluded that acupuncture was superior to various control interventions but there was insufficient evidence of its superiority to placebo (Ernst and White, 1998). Their conclusions appeared representative of the RCTs presented.

2.2.2.2 Summary of systematic reviews of manual therapy for LBP

Table 2.2 summarises the SRs I reviewed; further detail is provided within the narrative of this chapter.

Authors and Date	Title	Parameter dates	No. of Studies	Authors conclusions
Furlan, Giraldo, Baskwill, Irvin, Imamura, (2015)	Massage for low back pain	From database inception to August 2014	25 RCTs	Reported having very little confidence in the effectiveness of massage for LBP
Kumar, Beaton and Hughes, (2013)	The effectiveness of massage therapy for the treatment of nonspecific low back pain: a systematic review of systematic reviews	January 2000 to December 2012	9 SRs	Concluded that an emerging but very small evidence base existed, but indicated there was some evidence massage was effective for LBP in the short term
Hidalgo, Detrembleur, Hall, Mahaudens, Nielens, (2014)	The efficacy of manual therapy and exercise for different stages of non-specific low back pain: an update of systematic reviews	From January 2000 to April 2013	11 RCTs	Concluded manual therapy to be effective and favoured it as a treatment when combined with exercise
Slater, Ford, Richards, Taylor, Surkitt, Hahne, (2012)	The effectiveness of sub-group specific manual therapy for low back pain: a systematic review	From database inception to 2010	7 RCTs	Concluded that manual therapy was effective for the treatment of LBP when compared to other treatment
Rubinstein, Terwee, Assendelft, de Boer, van Tulder, (2012)	Spinal manipulative therapy for acute low-back pain	From 2000 to March 2011	20 RCTs	Concluded manipulation was no more effective than inert interventions, sham or other recommended therapies
Kuczynski, Schwieterman, Columber, Knupp, Shaub, Cook, (2012)	Effectiveness of physical therapist administered spinal manipulation for the treatment of low back pain: a systematic review of the literature	From database inception to May 2012	6 RCTs	Concluded spinal manipulation was safe delivered by physical therapists, the evidence supports its use, with improvement of clinical outcome
Assendelft, Morton, Yu, Suttorp, Sheklle, (2004)	Spinal Manipulative Therapy for Low Back Pain	From database inception to January 2000	39 RCTs and Q-RCTs	Determined no clinically relevant difference between spinal manipulation and other interventions for LBP

Table 2.2: Summary details of manual therapy SR's included in literature search

Furlan et al. (2015) conducted a SR of massage for LBP, reviewing databases to August 2014, (updating the 2008 SR of massage) they included 25 RCTs and the methods, results and meta-analysis presented were comprehensive.

The authors stated that the quality of the studies was low and reported having very little confidence in the effectiveness of massage for LBP. Their interpretation appeared representative of the RCTs presented.

A systematic review of SRs of massage for low back pain (Kumar, Beaton and Hughes, 2013) searched databases from January 2000 to December 2012 and included nine SRs. They reported significant methodological flaws in the primary research and concluded that an emerging but very small evidence base existed. A meta-analysis was not conducted. They indicated there was some evidence massage was effective for LBP in the short term. With the poor quality of some of the included SRs and primary research and the negative results of many of the SR, the conclusions of this SR appeared generous in nature.

Hidalgo et al. (2014) reviewed evidence for manual therapy from 2000 to 2013 and reported upon 11 RCT of medium to high quality with a low risk of bias. The methods and results were comprehensive, and they concluded manual therapy to be effective and favoured it as a treatment when combined with exercise. No meta-analysis was conducted but the conclusions appeared to be a fair representation of the studies included.

Slater et al. (2012) searched databases for manual therapy for LBP to 2010, and included seven studies. The methods and results were comprehensive, no meta-analysis was conducted, and they reported the quality of evidence was very low. They concluded that manual therapy was effective for the treatment of LBP when compared to other treatment. Their conclusions appeared not to consider the low quality of the RCTs and thus need interpreting with caution.

Rubinstein et al. (2012) updated their 2004 systematic review of spinal manipulative therapy for acute LBP in 2012, they included 20 RCTs, the quality of the studies was assessed using GRADE, one third of the RCTs were considered high methodological quality but the remaining studies were considered low and very low quality. The methods, results and meta-analysis of the SR were comprehensive, and they concluded spinal manipulative

therapy was no more effective than inert interventions, sham or other recommended therapies. The conclusions reached appeared to be representative of the RCTs included.

Kuczynski et al. (2012) conducted a SR of manipulation delivered by physical therapists (physiotherapists) they searched databases to May 2012 and included six RCTs for analysis. The methods and results of the SR were comprehensive, and they concluded the use of spinal manipulation was safe when delivered by physical therapists and the evidence supported its use, with improvement in clinical outcome. However they did note there was a lack of a true control group and a risk of bias across all the included studies. A meta-analysis was not conducted. The conclusions appeared to be a fair representation of the results presented.

The SR shown below (also see table 2.2) was also available in the March 2010 literature search and helped inform the planned pilot study for this thesis.

Assendelft et al's (2004) review of spinal manipulation for low-back pain searched databases up to January 2000, and 39 RCTs and quasi-randomised trials, with comparators of other manual therapies or no treatment were found. They reported that quality varied, but quality improved in more recent studies. The methods, results and meta-analysis were comprehensive, and they determined there was no clinically relevant difference between spinal manipulation and other interventions for LBP; they did not consider acupuncture in their review. The findings appeared representative of the RCTs presented.

2.2.2.3 Acupuncture versus manual therapy for LBP

One systematic review was found which attempted to compare exercise, acupuncture and manual therapy for LBP (Standaert et al. 2011, detailed in section 2.3.3), however it uncovered only a single publication for its review, which compared spinal manipulation with exercise for LBP. They concluded insufficient evidence was available in this area.

No systematic reviews or primary research were found combining acupuncture with manual therapy for LBP.

There were no protocols comparing or combining acupuncture with manual therapy for LBP registered in 2010, and this remained the case in 2018, with no new or current review protocols registered. While no protocols compared or combined acupuncture with manual therapy; one SR protocol of acupuncture only for (sub) acute non-specific LBP (Furlan et al. 2011) was found on the Cochrane database, and one SR protocol of manipulation only for LBP (Blanchette et al. 2015) was found on the BioMed Central (BMC) database.

Three individual studies (Giles et al. 2003; Cherkin et al. 2001; Giles et al. 1999) were identified from this literature review. The three studies compared acupuncture with manual therapy for LBP, and these studies all reported manual therapy to be favourable to acupuncture. These three studies are detailed in full in chapter seven, as part of the systematic review conducted for this thesis.

2.2.3 Evidence summary and conclusions

As detailed above, eight acupuncture SRs for LBP from 1998 to 2015 were reviewed. Two (Liu et al. 2015; Furlan et al. 2005) indicated acupuncture as an adjunct to conventional therapy was effective, and six (Liu et al. 2015; Zeng et al. 2015; Lam et al. 2013; Lee et al. 2013; Furlan et al. 2012; Ernst et al.1998) reported acupuncture to be effective when compared to usual care, medication or other interventions, but inconclusive when compared to sham. One (Furlan et al. 2005) concluded no firm conclusions could be drawn regarding the effectiveness of acupuncture for LBP (Furlan et al. 2005) and one reported insufficient evidence comparing acupuncture with spinal manipulation and exercise (Standaert et al. 2011).

Of the seven manual therapy SRs for LBP from 2004 to 2015 reviewed, two concluded manual therapy (Hidalgo et al. 2014: Slater et al. 2012) and one concluded manipulation (Kuczynski et al. 2012) were effective; two

(Rubinstein et al. 2012: Assendelft et al. 2004) reported manipulation was no more effective than other interventions or sham. One stated small but positive evidence for massage (Kumar et al. 2013) and one reported little confidence in the effectiveness of massage (Furlan et al. 2015).

The primary research across all the acupuncture and several of the manual therapy reviews (Furlan et al. 2015; Kumar et al. 2013; Slater et al. 2012; Rubinstein et al. 2012; Kuczynski et al. 2012; Assendelft et al. 2004) was reported as low quality with a high risk of bias, and significant methodological flaws. Only one manual therapy SR reported including medium to high quality studies with a low risk of bias in their review (Hidalgo et al. 2014). Meta-analyses were conducted in only four acupuncture (Lam et al. 2013; Furlan et al. 2012; Furlan et al. 2005; Ernst et al.1998), and three of the Manual therapy SRs (Furlan et al. 2015; Rubinstein et al. 2012; Assendelft et al. 2004), therefore the increased statistical power of combining results, and the improved estimates of effect sizes associated with meta-analyses, and thus greater confidence in the results, was not afforded to those SRs.

No SRs comparing acupuncture with manual therapy or combining the interventions were found in 2010. In 2018, one SR was found but as discussed, they concluded insufficient evidence existed (Standaert et al. 2011). Three individual studies (including one pilot study) comparing acupuncture with manual therapy were discovered in the initial 2010 search, these were conducted between 1999 and 2003. No further individual studies comparing the interventions were found in 2018.

No SRs or studies combining the acupuncture or manual therapy were found in 2010 or 2018 (other than the 2014 pilot study for this thesis).

At the time of my 2010 literature search only two acupuncture SRs (Furlan et al. 2005; Ernst et al. 1998) and one manual therapy SR (Assendelft et al. 2004) were available to inform the planning of the pilot RCT. Very limited evidence was available comparing acupuncture with manual therapy for LBP and no evidence available combining acupuncture with manual therapy for LBP. In

addition much of the evidence was based on studies with significant methodological limitations and risk of bias. Therefore from this review I determined there was a requirement for a further high quality RCT study into the effectiveness of acupuncture and manual therapy for LBP.

The additional research identified in 2018, after my 2010 review, remained limited in the comparison and combination of acupuncture and manual therapy for LBP, and the methodological and bias limitations persisted. Further research of high quality and rigorous methodological design was indicated in this area, to contribute to the knowledge base.

In the following sections of this chapter, I shall consider the international clinical practice guidelines, and how the available evidence for acupuncture and manual therapy was interpreted to inform the guidelines.

2.3 Clinical practice guidelines

Healthcare is a complex field with vast quantities of research of varying quality. Clinical guidelines are introduced and developed to provide guidance in the treatment of conditions, promote best evidence-based practice, reduce variation in treatment delivery, minimise the use of low-value interventions and thus improve patient care (O'Connell et al. 2017; Koes et al. 2010). They guide clinical practice, but are not necessarily intended to dictate clinical practice or to become regulatory for healthcare. Nevertheless they have had a tendency to become a prescription for best practice with one size fits all pathways, which can inevitably fail some individuals (Ault, 2018). However, NICE does refer to the responsibility of a treating practitioner:

> *"…the guideline does not override the individual responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient…"*

(NICE, 2016b, page 2; NICE, 2009c, page 1)

Early guidelines for LBP were less than popular amongst healthcare practitioners. In the USA the Agency for Health Care Policy and Research (AHCPR) produced a LBP guideline in 1994, which was opposed by powerful professional and commercial forces, to the detriment of the health service they intended to advise and support. Individuals demanded access to all interventions, as per the guideline and litigation suits and claims of sub-standard treatment ensued, a costly and lengthy process for the Health Service and Government (De Jong, 1995).

Delegated authority for guideline development in the USA was subsequently established and significant progress in the acceptance of guidelines ensued. Progress, in part, has occurred due to the involvement of healthcare professionals in the process, improvement and continually evolving guideline development process, peer review, improvement in the standards of guidelines, the quality and quantity of research, and advancements in medical interventions (WHO, 2014; Pillastrini et al. 2012; Bouwmeester, van Enst and van Tulder, 2009). Further efforts to support guideline acceptance remain a priority and efforts are continuing (Ault, 2018; O'Connell et al. 2017; Fischer et al. 2016; Glasgow et al. 2012).

Once developed, the implementation of guidelines is then required, which constitutes another hurdle. It is suggested that integrating recommendations from guidelines into clinical practice takes time, with the passage of time dependent upon the adaptations required by services and current clinical practice (NICE, 2016b). NICE (2016b) recommend aligning changes in recommendations to local priorities for the most effective outcome and transition of recommendations into clinical practice.

The process for disseminating clinical practice guideline recommendations is described as passive (Gad El-Rab, Zaïane and El-Hajj, 2017) with further support required for extracting information and facilitating it through to the clinical decision-making process needed (Gad El-Rab et al. 2017; O'Connell et al. 2017; Birch, Alraek, Lee, 2016; Ament et al. 2015).

A recent pilot survey of GP's concluded that their decision to offer access to acupuncture did not reflect their awareness of available guidelines; and reported acupuncture was most often accessed for pain relief for various musculoskeletal conditions including LBP. The study concluded that GP's often expected recommendations in guidelines when they were not in fact made (Czarnawska-Iliev et al. 2016).

As with all clinical practice, it is essential for healthcare professionals to take into account the most recent evidence, to inform their decision-making process in the delivery of care for LBP (O'Connell et al. 2017). However there remains uncertainty over the most appropriate management of LBP (Darlow et al. 2014). It would be anticipated that guidelines based on the same international evidence base would be similar, however recommendations can be conflicting (Koes et al. 2010). The production of over 39 different clinical guidelines for LBP, were reported internationally (Lonnemann et al. 2012).

Over time the general development and methodological quality of the guidelines has improved (Pillastrini et al. 2012), however there is a continued need to improve the development and quality of clinical guidelines (Pillastrini et al. 2012). It has been suggested that clinical practice guidelines are based on RCT evidence and expertise from solely wealthy countries and focus on treatment such as medication, and surgery and do not focus upon prevention and long term solutions. There is movement however for the redesign of treatment pathways and focus upon prevention to reduce LBPs impact and disability (Foster et al. 2018).

Healthcare professionals often assume consistency in recommendations across clinical practice guidelines. In section 2.4, I report my review of SRs of international clinical practice guidelines and consider the similarities and differences across the published international guidelines.

2.4 Systematic Reviews of International Clinical Practice Guidelines

With the existence of many international guidelines there have subsequently been systematic reviews comparing clinical practice guidelines for low back pain and these are discussed below.

Prior to reviewing the international clinical practice guidelines, I conducted a search of available SRs of clinical practice guidelines to appraise the current evidence.

2.4.1 Aims and methods

Aim of the review:

- *To review systematic reviews of clinical practice guidelines for LBP*

In 2018, I conducted electronic database and internet searches for systematic reviews of clinical practice guidelines, for the period of 2000 to 2017. The SRs needed to include LBP guidelines, including information on the diagnosis and treatment of LBP. I extracted data regarding the included guidelines, the diagnosis and treatment of LBP and the methods of the SRs.

2.4.2 Results of SRs of clinical practice guidelines

Three SRs were identified and included in my review and are detailed below.

Koes et al. (2001, 2010): In 2001, a review included the clinical practice guidelines from 11 countries, published between 1994 and 2000. One guideline for each country was included; if multiple updates had occurred the most recent publication was included. Systematic review methods were used, using electronic databases and the methods were comprehensive and appeared robust. Four authors were used to extract data from three or four guidelines, they reported using the same data categories of their 2001 review,

however limited detail was provided and no information regarding consistency or any disparities between the authors extraction was given, limiting the transparency of the review.

The review established that similar recommendations were found between the diagnostic classification and therapeutic interventions. However, discrepancies between recommendations on exercise therapy, spinal manipulation, muscle relaxants and patient information were found (Koes et al. 2001).

In 2010, the 2001 systematic review was updated and included 13 national clinical guidelines and two international clinical guidelines from Europe, (one for acute LBP and one for chronic LBP) (Koes et al. 2010). The search extended the previous review from 2000 to 2008 and guidelines from the 2001 review were also checked for updates. One guideline was selected for each country; the selection was based upon the most recent guidelines published. Table: 2.3 lists the guidelines included within the 2010 systematic review.

Country	Title	Author
Australia, 2003	Australian acute musculoskeletal pain guidelines. Evidence-based management of acute musculoskeletal pain.	Bowen Hills: Australian Academic Press; 2003
Austria, 2007	Evidence and consensus based Austrian guidelines for management of acute and chronic nonspecific backache	Friedrich, M. Likar, R. 2007
Canada, 2007	An interdisciplinary guideline development process: The Clinic on Low-back pain in interdisciplinary Practice (CLIP) low-back pain guidelines	Rossignol et al, 2007
European guidelines, 2004	European guidelines for the management of acute nonspecific low back pain in primary care	van Tulder et al, 2006
European guidelines, 2004	European guidelines for the management of chronic nonspecific low back pain	Airaksinen et al, 2006
Finland, 2008	Low back pain among adults. An update within the Finnish current care guidelines	Malmivaara et al, 2008
France, 2000	Guidelines department, diagnosis and management of acute low back pain (<3 months) with or without sciatica & diagnosis, management and follow-up of patients with chronic low back pain	Agence Nationale d'Accreditation et d'Evaluation on Sante (Agence, 2000)
Germany, 2007	Recommendations for treatment of low back pain	Drug, 2007
Italy, 2006	Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines	Negrini et al, 2006
New Zealand, 2004	New Zealand Acute Low back pain Guide	NHC, 2004
Norway, 2007	New clinical guidelines for low back pain	Laerum et al, 2007
Spain, 2005	Lumbalgia Inespecifica. Version espnola de la Guia de Practica Clinica del Programa Europeo COST B13	Spanish Back Pain Research Network (2005)
The Netherlands, 2003	Clinical guideline for non-specific low back pain	Dutch Institute for Healthcare Improvement (CBO) 2003
United Kingdom, 2008	Back pain (low) and sciatica	Clinical knowledge summaries (CKS) NICE, 2008
United States, 2007	Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society	Chou et al, 2007

Table 2.3: Systematic review of clinical guidelines 2010 (Koes et al. 2010)

Koes et al. (2010) reported their expectation was to discover similar guideline recommendations, as the guidelines were all based on the same international

evidence. Similar recommendations remained in place from the 2001 review, with specific diagnostic triage and classification, following a red flag diagnostic process and physical examination (though European guidelines advocate a limited neurological screen comparative to the other guidelines). Diagnostic imaging was not recommended for routine cases in any guidelines. Recognition of psychosocial factors (yellow flag systems) was seen across all the guidelines, though variation in the detail of assessing these factors was seen. Table 2.4 summarises the common diagnostic recommendations:

Common Recommendations for Diagnosis of Low back pain
Diagnostic triage (non-specific low back pain, radicular syndrome, serious pathology)
Screen for serious pathology using red flags
Physical examination for neurological screening (including straight leg raise test)
Consider psychosocial factors (yellow flags) if no improvement
Routine imaging not indicated for non-specific low back pain

Table 2.4: Summary of recommendations for diagnosis of LBP (Koes et al. 2010)

Therapeutic interventions recommended are summarised in table 2.5, across all the guidelines these included the discouragement of bed rest, the recommendation of multidisciplinary team approaches, chronic pain supervised exercise programmes and cognitive behavioural therapy.

Though significant consistency is seen across the international guidelines, variations were reported. Discrepancies concerning prescription medication remain across the guidelines. Paracetamol was generally recommended along with non-steroidal anti-inflammatories, however vast differences were seen for the recommendation of opioids, muscle relaxants, benzodiazepines and anti-depressants despite the evidence base being consistent. It is suggested that cultural norms of medication use in each country may explain the

variation (Koes et al. 2010). Spinal manipulation remained variable from the previous review in 2001. Across the guidelines some recommended manipulation or advised it as a short-term therapeutic option, and others did not recommend it. This was attributed to evidence not being robust enough for consistent recommendation, allowing interpretational differences at guideline committees (Koes et al. 2010). Koes et al. (2010) noted the differences were less than may be anticipated when considering varying cultures and healthcare systems, and they attributed the differences to disparities in local care systems and lack of robust evidence for some interventions.

Common Recommendations for the Treatment of Low back pain	
Acute or Sub-acute Pain	**Chronic Pain**
Reassure patient (favourable prognosis)	Ultrasound & electrotherapy - not recommended
Recommended to remain active	Medication – short term use
Medication as indicated	Manipulation – short term use
Bed rest discouraged	Supervised exercise therapy - recommended
Supervised exercise programmes – not recommended	Cognitive behavioural therapy - recommended
	Multidisciplinary approach - recommended

Table 2.5: Summary of recommendations for the treatment of LBP (Koes et al. 2010)

The included guidelines were based on scientific evidence, consensus from committee discussion and expert opinion. However the guideline reports did not disclose detail of which decisions were influenced by committee consensus or expert opinion and did not publish the committee or expert discussions, limiting transparency of the guidelines in the review. Additionally, limited focus on whether cost effectiveness data influenced decisions was not reported across the guidelines (Koes et al. 2010).

The Koes et al. 2010 update established there had been minimal changes to the guideline recommendations a decade on. The key developments included no longer recommending MRI and CT scans for routine cases and taking a more robust approach in the consideration of psychosocial factors to prevent chronicity. Therapeutically, exercise therapy held a firmer position in the recommendations, along with remaining active and returning to work as soon as possible. Some clarity regarding primary and secondary medication approaches and when to refer to primary or secondary care were provided but overall the management of LBP since the 2001 review remained consistent. It may be that LBP was being managed appropriately as per the earlier guidelines, or that sufficient new evidence had not been established for amendments to be made to the recommendations (Koes et al. 2010).

Koes et al (2010) reported all the included guidelines were based on a comprehensive literature reviews (Cochrane library, Medline, Embase, PEDro) using systematic review methods, with most guidelines adopting a weighting of the strength of evidence and rating of evidence to determine their recommendations. Three committees (Austria, Germany and Spain) based their evidence entirely or in part on the European LBP guidelines. Previous guidelines and literature reviews are reported as more frequently becoming the starting point for new searches and guidelines (Koes et al. 2010).

The 2010 review by Koes et al. (2010) presented only national multidisciplinary guidelines and did not include an exhaustive list of all clinical guidelines, and many guidelines therefore were not included within this review. This review was an update of the 2001 review however no planned follow up of the review is reported. Koes et al. (2010) acknowledged this as a limitation of their review. An additional limitation transpires due to all the included guidelines originating from wealthy industrialised countries limiting the reviews global generalisability.

Pillastrini et al. (2012): Pillastrini et al. (2012) conducted a systematic review including 13 clinical practice guidelines for LBP from eight separate countries and one European guideline. SR methods were used, and the methods appeared explicit and robust. Electronic databases were searched

for clinical practice guidelines from 2002 to 2010. Table 2.6 shows the included guidelines.

Country	Title	Author
United States, 2005	Institute for clinical systems improvement (ICSI): United States	ICSI, 2005
United States, 2007	Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians (ACP); American Pain Society low back pain guidelines panel (APS)	Chou et al, 2007
Canada, 2009	Guidelines for the evidence – informed primary care management of low back pain	Institute of Health Economics, towards optimized practice (TOP) program, 2009
Canada, 2007	Clinic on low back pain in interdisciplinary practice (CLIP) guidelines;	Rossignol et al, 2007
Australia, 2002	Low Back Pain	New South Wales Therapeutic Assessment Group Inc (NSW TAG) 2002
European guidelines, 2004	European guidelines for the management of chronic nonspecific low back pain	Airaksinen et al, 2006
Italy, 2006	Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines	Negrini et al, 2006
The Netherlands, 2003	National practice guideline for the physiotherapeutic management of patients with low back pain	Bekkering et al. 2003
United Kingdom, 2009	Low back pain: early management of persistent non-specific low back pain. National Collaborating Centre for Primary Care and Royal College of General Practitioners	Savigny et al, 2009
United Kingdom, 2006	Clinical guidelines for the physiotherapy management of persistent low back pain. Chartered Society of Physiotherapy	Mercer et al. 2006
Austria, 2007	Centre of Excellence for Orthopeadic Pain Management Speising (CEOPS)	Friedrich & Likar, 2007
Germany, 2007	Recommendations for the treatment of low back pain.	Drug Committee of the German Medical Society, 2007
Germany, 2003	The German College of General Practitioners and Family Physicians (DEGAM) low back pain guideline	Becker et al. 2003

Table 2.6: Systematic review of clinical guidelines 2012 (Pillastrini et al. 2012)

The guidelines selected by Pillastrini et al (2012) in their review differed from those included by Koes et al. (2010) and appeared to focus more on specific

professional association guidelines as opposed to national guidelines, for example they included the Chartered Society of Physiotherapy Clinical Practice guidelines (CSP, 2007). It was also the first review to include interventions such as acupuncture, hydrotherapy etc.

As part of the review, three authors assessed the methodological quality of the identified clinical guidelines independently, using the Appraisal of Guidelines for Research and Evaluation (AGREE) tool. 13 guidelines were reported to be satisfactory, using effective literature review methods, scope and clarity, with adequate critical appraisal of the evidence and thus were included within the review.

This review reported a strong consensus across the guidelines regarding diagnostic triage and prognostic factors, and an increased alignment of therapeutic interventions; which included exercise therapy, multidisciplinary approaches and combined physical and psychological interventions. An improvement in the quality of chronic pain guidelines was noted in this review (Pillastrini et al, 2012).

This review included guidelines which:

- *Addressed the clinical management of non-specific chronic LBP in primary care*
- *Were published by a professional group*
- *Were available in English, Italian or German languages*
- *Stated recommendations for therapeutic interventions explicitly*

(Pillastrini et al. 2012)

It is of interest that this review opted for guidelines published by professional groups, as this potentially may have excluded National guidelines if they were produced by a guideline development committee and not produced by a professional body such as the Chartered Society of Physiotherapy, or the Royal College of General Practitioners. It may have been they were attempting to focus primarily upon clinical interventions delivered by

healthcare professionals, but Pallastrini et al. (2012) do not offer any rational or discussion to allow for understanding of their selection criteria.

Pillastrini et al. (2012) determined the diagnostic recommendations continued to be consistent with Koes et al. (2010) findings, with no differences found, he also clarified the unified approach across the guidelines of focus upon identifying risk factors to prevent chronicity.

Therapeutic recommendations were reported as more varied across the guidelines and this was attributed to the significant involvement of multiple disciplines in the management of LBP (Pillastrini et al. 2012) table 2.7 summarises the key recommendations.

Recommendations for the Treatment of Low back pain
Educational interventions - recommended
Manipulation – recommended or therapeutic option or not recommended
Supervised exercise therapy - recommended
Cognitive behavioural therapy - recommended
Multidisciplinary approach - recommended
Tens, biofeedback, lumbar supports, ultrasound, electro, laser therapy and traction – not recommended

Table 2.7: Summary of recommendations for the treatment of LBP (Pillastrini et al. 2012)

Pillastrini et al. (2012) discussed that although there was an array of conservative interventions available from physiotherapists, few were routinely recommended across the guidelines due to a varied evidence base. A strong consensus and evidence for exercise interventions was reported but advised a lack of clarity in type of or prescription of such exercises across the guidelines. Consensus also existed in recommending combined physical and psychological interventions across the guidelines (Pillastrini et al. 2012). Spinal manipulation continued to not be consistently recommended across the guidelines, which Pillastrini et al. (2012) attributed to the continued

underlying conflicting evidence and the interpretation from the guideline development committees, he endorsed the need for future consensus on precise recommendations for manual therapy.

Limiting their review to professional group publications only limited the scope of this review and thus the inclusion of the most current guidelines was not included within the review. Additionally, two different clinical guidelines from the USA, Canada, the UK and Germany were included in the review, questioning the breadth of the review and limiting the international generalisability and detail of the review. Pillastrini et al. (2012) offered no insight into their methodological reasoning and identified no limitations of their review. They focused solely on the limitations and recommendations of the included guidelines. Pillastrini et al. (2012) recommended future guidelines should adopt the GRADE framework to grade evidence and inform recommendations (GRADE, 2004) to limit confusion and allow international consistency and collaboration. No planned review date was reported within the review.

Wong et al. (2017): More recently, Wong et al. (2017) conducted a systematic review of clinical practice guidelines for LBP. Systematic review methods were used, and the methods were comprehensive and appeared rigorous. They searched electronic databases from 2005 to 2014.

13 guidelines were discovered from five different countries and two European guidelines. The systematic review included more than one guideline from each country and also included multiple editions of guidelines (Wong et al. 2017) thus limiting the geographical scope of the review.

The guidelines were critically appraised, using the AGREE II criteria by independent reviewers. Ten of the clinical practice guidelines were selected and recorded as having a low risk of bias and high methodological quality, nine of the ten high quality guidelines focused of low back pain (without diagnostic specificity) and were included within the review. The tenth high quality guideline was specific to lumbar disc herniation with radiculopathy; this guideline was not included in the review due to its condition specificity

to one type of LBP (Wong et al. 2017). All included guidelines used robust literature review methods and adequate criteria for critically appraising of the evidence (Wong et al. 2017). Table 2.8 shows the guidelines included in the review.

Country	Title	Author
European guidelines, 2004	European guidelines for the management of chronic nonspecific low back pain	Airaksinen et al. 2006
European guidelines, 2004	European guidelines for the management of acute nonspecific low back pain in primary care	van Tulder et al. 2006
Belgium 2006	Chronic low back pain (KCE report)	Nielens, 2006
United States, 2007	Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians (ACP); American Pain Society low back pain guidelines panel (APS)	Chou et al. 2007
United States, 2009	An evidence based clinical practice guideline from the American Pain society	Chou et al. 2009
United Kingdom, 2009	Low Back Pain: Early management of persistent non-specific LBP	NICE, 2009
United States, 2011	A joint clinical practice guideline from the American College of Physicians and the American Pain Society	Livingston et al. 2011
Canada, 2011	Alberta health technology assessment (HTA) Canada: Multidisciplinary clinical practice guideline low back pain	Cutforth et al. 2011
Scotland, 2013	Scottish intercollegiate guidelines network (SIGN) Management of chronic pain	SIGN, 2013

Table 2.8: Systematic review of clinical guidelines 2017 (Wong et al. 2017)

Wong et al (2017) excluded guidelines if they were deemed to have poorly conducted systematic literature review searches or if inadequate critical appraisal methods of the evidence were evident. The three low methodological quality guidelines identified included a guideline from the Phillippines (PARM, 2011), the American physical therapy guideline (Delitto et al. 2012) and an Ottawa clinical practice guideline (Brosseau et al. 2012),

and these were excluded from the review for not fulfilling the quality standards of the review (Wong et al. 2017).

In this review, individual countries within Europe appear to have been omitted with the exception of the UK and Belgium. The European guidelines were however included.

The authors concluded that some methodological improvements were apparent from earlier reviews, but suggested updates were required to many guidelines, and limitations existed across the guidelines, notably in guideline applicability, (including facilitators and barriers), resource implication and monitoring criteria upon implementation were reported as not being adequately addressed. They reported evidence of consensus in recommending education, activity and exercise, manual therapy and paracetamol or NSAIDs as first line treatments for LBP (Wong et al. 2017). Wong et al. (2017) recommended many guideline developers needed to further address the implementation of the guideline recommendations and also the involvement of the end user in the development of future guidelines.

Wong et al. (2017) reported some limitations of their review, which included selecting only English language guidelines (limiting the external validity). They also reported the differing use of LBP classifications (i.e. acute, sub-acute, chronic) limited their ability to ensure misclassification did not occur in their synthesis. No planned review date was reported within the review.

2.4.3 Summary of the systematic reviews:

Allowing for differences in the dates completed, it could be anticipated that the SRs of the clinical practice guidelines for LBP would consider similar guidelines and countries. However across the three SRs presented (Koes et al. 2010, Pillastrini et al. 2012, Wong et al. 2017) only two clinical practice guidelines (United States, Chou et al. 2007 and European, Airaksinen et al. 2006) were common across all three SRs. Recommendations did appear similar with some changes in the treatments recommended over time.

Across the various guidelines the recommendations of acupuncture and manual therapy were inconsistent, section 2.5 reviews the international clinical practice guidelines and considers the recommendations for acupuncture and manual therapy and the inconsistencies across the guidelines.

2.5 International Guidelines – Recommendations for Acupuncture and Manual therapy

2.5.1 Indication for a review of international clinical practice guidelines, recommendation of acupuncture and manual therapy

Various countries have developed national clinical practice guidelines in attempt to manage LBP. As all the guidelines are based upon the same international clinical evidence, it would be anticipated that the guidelines would provide similar recommendations for diagnosis and treatment interventions. The review published in 2010 reported the guidelines to be similar regarding diagnostic classification and therapeutic interventions, however significant differences existed across countries (Koes et al. 2010).

The guidelines and systematic reviews of guidelines discussed above debate the evidence for manual therapy but there is minimal discussion or acknowledgement of acupuncture, when it could be reasoned that the evidence for acupuncture is no less controversial than the evidence for manual therapy.

While RCTs of acupuncture have generally demonstrated poor quality (Li et al. 2016), there are several RCTs assessing the efficacy of acupuncture, which have been assessed as having high methodical quality using the Cochrane Collaboration Back Review Group (CBRG) quality assessment for RCT's. These studies have reported that acupuncture can significantly reduce pain intensity in patients with LBP (Liang et al. 2017; Brinkhaus et al. 2006; Thomas et al. 2006; Leibing et al. 2002; Molsberger et al. 2002).

A recent meta-analysis of chronic pain demonstrated acupuncture to be effective at reducing chronic pain (Vickers et al. 2018b) and it was also shown to have lasting effects beyond 12 months for relieving pain (MacPherson et al. 2017b). A Cochrane review in 2005 concluded that acupuncture was more effective for pain relief than no treatment or sham acupuncture for the treatment of chronic LBP at a three-month follow up. While acupuncture was not shown to be more effective than other conventional therapies; when it was added to other conventional therapies, it was shown to improve pain and function better than conventional therapies alone (Furlan et al. 2005; Dascanio et al. 2015e).

There is limited high quality evidence for the efficacy of acupuncture for the treatment of LBP, many publications include case studies / series and intervention studies with methodological designs that are inadequate for assessing efficacy. Lewis and Abdi (2010) suggested the quality of studies conducted in more recent years have improved; however effectiveness of acupuncture remains to be conclusively demonstrated. The available evidence for acupuncture is discussed further in section 2.7.2 of this chapter.

2.5.2 Aims and methods

Aim of the review:

- *To review the clinical practice guidelines for LBP and their recommendations of acupuncture and manual therapy.*

In 2018, I conducted electronic database and internet searches for international clinical practice guidelines, for the period of 2000 to 2017. I also searched the reference lists and included guidelines in the SRs detailed in section 2.4, and checked the reference lists of any guidelines found. For inclusion in the review, the guidelines needed to be national guidelines for LBP, and include information on the diagnosis and treatment of LBP. Only one guideline from each country was included, the most current guideline was reported upon, however the previous edition of the guidelines was considered

to evaluate any changes which may have occurred in the recommendations over time. I extracted data regarding the methods of the guidelines and the recommendations of acupuncture and manual therapy.

2.5.3 Results of SRs of clinical practice guidelines

Clinical practice guidelines from eight countries and one European guideline were identified and included in my review and are detailed below.

USA: The American College of Physicians (ACP) produced an updated guideline in 2017 (Qaseem et al. 2017). The updated guideline focused upon non-invasive pharmacologic and non-pharmacologic treatments of LBP. The target population included adults (>18 years) with acute, sub-acute and chronic LBP, radicular LBP or symptomatic spinal stenosis. Published studies between January 2008 and November 2016 were included along with earlier studies identified by their inclusion in the 2007 American Pain Societies systematic review.

The ACP policy also detailed if guidelines were not updated within five years they are automatically withdrawn, ensuring their guidelines are always current and up to date (Qaseem et al. 2010). The development process for this guideline appeared to be fully transparent and robust, thus providing confidence in its recommendations.

Qaseem et al. (2017) determined the guideline development process to be of high quality, it was developed and established by the American College of Physicians (ACP) and the methods of the development process were fully transparent, (summarised in the guideline, included fully as an appendix of the guideline and also published separately). The guideline development followed a multistep process, included a systematic review methodology of the evidence base, deliberation of the evidence by the committee, summary recommendations with evidence and recommendation grading (Qaseem et al. 2010). The guideline underwent a peer review process and a consultation process (including public) prior to the final guideline publication (Qaseem et

al. 2017). Throughout the guideline the recommendations are linked to the evidence and the evidence is graded, with the strength of the guideline recommendation also indicated.

The most significant modification in the updated American guideline publication was that the guidelines recommended first line treatment should consider non-drug therapy, in an attempt to move away from reliance upon pharmaceutical treatments for LBP. For chronic LBP they recommended:

> "exercise, multidisciplinary rehabilitation, acupuncture, mindfulness-based stress reduction (moderate quality evidence) tai chi, yoga, motor control exercise, progressive relaxation, electromyography biofeedback, or spinal manipulation (low-quality evidence)."

(Hackethal, 2017, page 3)

For chronic LBP that was non-responsive to the above non-drug therapy interventions, NSAIDS were recommended as a first line medication. This was followed by tramadol, duxetine, with opioids only considered if first and second line drug therapy was ineffective. Not offering drug therapy, as first line treatment was a considerable shift in the approach within primary care internationally. However it was suggested that the approach was due to limited choices of pharmacology, and to the progression in approach aiming to prevent LBP becoming a chronic condition (Atlas, 2017).

In their conclusions, the ACP summarised that there was no evidence to suggest one therapy was clearly better than any other, though more recent evidence supported mindfulness-based stress reduction techniques, tai chi for chronic LBP and acupuncture for acute LBP (Qaseem et al. 2017).

Canada: The Institute of Health Economics developed the clinical practice guidelines for LBP with a multidisciplinary team of clinicians and researchers: Towards Optimized Practice (TOP) program in 2015 (TOP, 2015) and updated with minor revisions in 2017 (TOP, 2017). The guideline focused upon diagnosis and conservative non-surgical interventions for LBP.

The target population included adults (>18 years) with LBP of any duration (excluding pregnant women).

To ensure the guidelines remain current, a scoping search of the literature to inform the guidelines is conducted annually to assess if the guidelines require updating, if two or more new good quality SEED (clinical practice) guidelines were identified then an update of the guidelines was commenced. This maintained the currency and accuracy of the guideline, in line with the best available evidence.

The Canadian guidelines appeared to be of high quality with comprehensive methods and reporting, using systematic review methods. The methods of the Canadian guidelines appeared comprehensive with a background document, supporting documents and a process description document, all published as supportive documentation. Two committees were involved, a Steering Committee (SC) tasked with collection and collation of research and operational oversight, and the Guideline Update Committee (GUC) who reviewed the 2nd edition of the guideline and revised the recommendations in the 3rd edition to reflect the current research in LBP management. The guidelines were peer reviewed by various stakeholders, professionals with experience and interest in pain management, their colleagues, and patients with acute and chronic LBP.

Decisions and recommendations were made based on the collective professional opinion of the committees and analysis of the relevant evidence. Limited information was provided regarding the professional opinion of the guideline development committee, while they state expert opinion was used when evidence was older the seven years, no information is provided to support the discussions or opinions. It was not possible to clarify if recommendations were influenced by professional opinion or the weight given to the professional opinion, comparative to the evidence. Potentially allowing professional bias to influence the guideline recommendations and thus they lacked transparency in this area.

The Canadian guideline relied profoundly upon SEED guidelines from other countries to inform its guideline. No reviews of original evidence or studies were conducted for this guideline. Thus the guideline committees did not assess the quality of the empirical evidence directly, and any recently published evidence was not necessarily incorporated into the guideline, limiting the opinion and the strength of the information provided.

The SEED guidelines were critically appraised on their methodology and reporting using the AGREE tool (Appraisal of Guidelines for Research Evaluation, (AGREE, 2009) and only those rated as average or good were included. SRs of new interventions not covered by the previous Canadian guidelines, or other SEED guidelines were critically appraised by the GUC, though no detail is provided.

Throughout the guideline however evidence for individual recommendations were referenced to other guidelines and evidence, i.e. acupuncture for chronic pain references supporting evidence of SRs from the previous Canadian guidelines and the Scottish guidelines. A comprehensive evidence list for each individual condition was not provided in the updated guidelines, limiting the information provided.

In similarity to the America clinical practice guideline, the Canadian guideline recommended both acupuncture and manual therapy for LBP. However the Canadian guidelines recommend manual therapy (spinal manipulation) for acute LBP, but did not recommend for or against acupuncture or massage due to insufficient evidence. For chronic LBP the guideline recommends acupuncture, recommends massage but did not recommend for or against spinal manipulation due to insufficient evidence (TOP, 2017). Therefore disparities in the recommendations of acupuncture and manual therapy exist between the American and Canadian guidelines.

Europe: The European guidelines for the management of chronic low back pain were published in 2006 (Airaksinen et al. 2006). The guideline focused upon the diagnosis, treatment and management of chronic LBP across Europe; an age demographic was not specified for the guideline target

population. The guidelines were developed to inform groups who were intending to develop new or review current guidelines, indirectly informing the public, health care providers, industry and policy makers in Europe.

The guidelines have not been updated since 2006 and thus not current in there evidence base. No information is afforded regarding planned reviews or subsequent editions of the guidelines, and this is a significant limitation of the European guideline, which was intended to inform the development of other national clinical practice guidelines.

The guideline Working Committee (WC) consisted of experts in the field of LBP research and developed the guideline within the COST ACTION B13 (2006) framework LBP guidelines issued by the European Commission. A draft guideline was produced and reviewed by members of the management committee of COST B13 before the full guideline was published. While drafts were circulated to the committee and the working group, no external peer review or public consultation of the document occurred, limiting its editorial independence and transparency.

The guideline used systematic reviews of SRs, RCTs and searched existing national guidelines. The quality of evidence was assessed for systematic reviews using the Oxman and Guyatt index (Oxman and Guyatt, 1991) and for RCTs a methodological quality assessment, with criteria related to internal validity (van Tulder et al. 1997). The guideline rated evidence from A (strong evidence) to D (no evidence) and their recommendations were based on evidence with a quality rating of A or B. The quality assessment techniques used for the European guidelines were outdated and may not be as informative as more recent assessment criteria.

Committee discussion was used to consider the evidence and inform the recommendations though no information was provided outlining the discussions or their influence on the recommendations, limiting the transparency of this process and subjecting the guidelines to potential bias from committee members.

The guideline concluded no single intervention is likely to be effective in the treatment of LBP and that effect sizes of therapeutic interventions were rather modest. The guidelines considered the available evidence for acupuncture, but were unable to recommend it as a treatment option, due to the conflicting evidence surrounding acupuncture and sham acupuncture and limited evidence when compared to other interventions. While they acknowledged the evidence for manual therapy (including spinal manipulation) was not conclusive due to the low quality of many studies within systematic reviews, they concluded they were able to recommend a short course of manipulation / mobilisation as a treatment option for chronic LBP. They were unable to recommend massage due to insufficient evidence (Airaksinen et al. 2006).

Belgium: The guidelines in Belgium were published in 2017 (Van Wambeke et al. 2017) updated from their previous guideline in 2006 (Nielens et al. 2006), by the KCE (Belgium Health Care Knowledge Centre) an organisation part funded by the public service. The guideline focused on the evaluation and management of LBP. An age demographic was not specified for the guideline target population, however due to them following the NICE guideline they do refer to the NICE target population of those >16 years with LBP of any duration.

No guideline update or review was conducted of the 2006 guideline until the publication of the 2017 guideline, however in their current guideline they recommend the guideline would ideally be reviewed at five years following publication to ensure it is current. However no process is documented to outline how any future review process will occur.

The Guideline Development Group (GDG) consisted of multidisciplinary healthcare professionals and KCE researchers. The guideline development process appeared to be robust and of high quality, it was developed and published as a separate document and was fully transparent. The guideline was produced following a standard methodology based on a systematic review of the evidence and they searched for previous clinical practice

guidelines. All clinical practice guidelines were appraised using the AGREE II checklist (AGREE, 2017).

The KCE guidance noted that their recommendations were based largely on the NICE guidelines (2016) which they reviewed in depth and referred specifically to them throughout as part of their guideline development, however they did not follow all the recommendations made by NICE, outlining any differences and justifications for their decisions. Similarly in there 2006 guideline where they referred to the European Clinical Practice Guidelines (Airaksinen et al. 2006) as an important source of evidence for its guideline but did not follow all its recommendations. This approach limited the GDGs assessment of the source data and published evidence.

The guideline recommendations were circulated to professional associations and clinical experts for external peer review, though no open review or public consultation was conducted, limiting the scope of the external review. Comments were gathered and discussed before the final external validation of the guideline, by two clinicians and the methods reviewed against the AGREE II checklist.

The published guideline clearly indicated the levels of evidence for each recommendation and if expert opinion was used. While it is useful to know the basis for the recommendations, no information on the expert opinion or discussions is detailed regarding their influence on the recommendations, limiting the transparency of the recommendations and the potential for committee bias to be introduced.

Recommendations in the 2017 guideline were similar to the 2006 guideline (except for acupuncture and medication) in which they concluded remaining active and minimal time missed from work is essential. They favoured conservative treatments with moderate and high-quality evidence over surgery or invasive procedures and recommended both acupuncture and manual therapy. The guidelines also recommended against bed rest, lumbar supports and paracetamol (Van Wambeke et al, 2017; Nielens et al. 2006). They reported low quality evidence was available to recommend massage,

TENS, and hydrotherapy. Moderate quality evidence was reported for recommending Back schools, educational programmes for self-care, spinal manipulative / mobilisation therapy (and acupuncture in 2006). High quality evidence was reported for recommending the modest effectiveness of exercise programmes, multi-disciplinary bio-psychosocial rehabilitation, CBT (Van Wambeke et al, 2017; Nielens et al. 2006).

In their updated 2017 guidelines they recommended considering manipulation, mobilisation and soft-tissue techniques but only as part of a multimodal treatment with supervised exercise, the level of evidence ranged from high to very low, with the strength of the GDG recommendation being weak. For acupuncture the guideline committee reported a recommendation could not be formulated, due to the conflicting evidence and the recent change in the NICE guideline recommendations. They did not follow the NICE guideline by not recommending it but instead took a neutral position, stating 'No recommendation on acupuncture has been formulated', this was a change from their 2006 guideline in which they recommended both acupuncture and manual therapy (Van Wambeke et al, 2017; Nielens et al. 2006).

Netherlands: The KNGF (Koninklijk Nederlands Genootschap voor Fysiotherapie) produced the Dutch clinical practice guideline for patients with low back pain in 2013 (KNGF, 2013), updating its previous guidelines from 2003. An age demographic was not specified for the guideline target population.

No information is afforded regarding the process for updating the guidelines. The Dutch college of general practitioners also produced a guideline for LBP however this dates back to 1996 with no recent updates. The KNGF guideline was focused upon diagnostics and treatments of patients with non-specific LBP by a physical or manual therapist.

The guideline followed SR methods, but limited information regarding the methodological process of developing the guidelines was available limiting its transparency. However Pillastrini et al. (2012) considered it to be a high-quality guideline, considering evidence from systematic reviews and meta-

analyses. No information regarding the quality of assessment of evidence was provided and no direct links were made to the evidence in their recommendations, limiting the depth and critical information in the guideline.

A multidisciplinary committee was formed to develop the guidelines and their opinion was used within the guidelines, though no information regarding the weight of professional opinion to the evidence was provided limiting the transparency of the recommendations and potentially allowing for bias from committee members. No external peer review or public consultation was conducted in the production of these guidelines, limiting the editorial independence of the review.

The primary advice for the KNGF guideline was patient reassurance and maintaining and increasing activity levels, and they recommended the consideration of manual therapy (joint mobilisation, manipulation or massage) (KNGF, 2013). However Acupuncture was not considered within the Dutch physiotherapy guidelines in 2003 or 2013, this was potentially due to the guidelines being produced by the Dutch physiotherapy association and physiotherapists not being permitted to practice acupuncture in their country, unless they have completed an additional full-time three-year degree in acupuncture and thus acupuncture is not commonly practiced by physiotherapists in primary care.

France: The French guidelines were published in 2016 (Petit et al. 2016) and focused on the prevention of low back disorders in the working population, though an age demographic was not specified for the guideline target population.

The guideline development committee consisted of a multidisciplinary team of 24 experts. The guideline methods followed the Clinical Practice guideline methods proposed by the French National Health Authority. SR methods were followed, and the evidence was assessed providing a grade for strength of each recommendation in relation to the evidence or expert opinion. Clear indication of when expert opinion was used in recommendations was given though no detail regarding the discussions or opinions is available, limiting

the transparency. A committee of 50 experts completed the peer review, though no external peer review or public consultation was conducted, limiting the editorial independence of the review.

The French guidelines advised there was no evidence to recommend one form of manual therapy over another but recommended manual therapy, in line with the European guidelines. They reported being unable to recommend acupuncture due to conflicting evidence and lack of strong evidence to support the intervention (Petit et al. 2016).

Italy: The Italian guidelines were financed by the Italian Health Ministry and published in 2006 (Negrini et al. 2006). They focused upon the diagnosis and therapeutic input for LBP of any duration, and produced flow diagrams for a variety of LBP conditions. Age demographic was not specified for the guideline target population.

The guidelines state a review and a further literature search would be anticipated by 2008, however no evidence of a review of the guidelines is apparent, thus the guideline does not include the current evidence base due to its publication in 2006.

A committee of multidisciplinary health professionals developed the guideline, and an epidemiologist performed the 'bibliographic research', evaluated the methodological quality and synthesised the data into tables for the committee. No information regarding the epidemiologist is provided to determine their expertise in the various areas of LBP research. No information regarding peer review or public consultation is evident, limiting the editorial independence of the guideline.

SR methods were used, they considered SRs, RCTs and clinical practice guidelines, though limited information regarding the methods was provided. An epidemiologist conducted a 'critical methodology quality evaluation' of the literature, though the quality evaluation process is not transparent within the guideline. Evidence was graded as 'A' a strong recommendation, through to 'C' there is deep uncertainty pro or versus the recommendation. They

reported their grading system did not conform strictly to the levels of evidence, thus limiting the interpretation of the evidence grading. Additionally while an evidence rating is given there is no corresponding evidence linked to the guideline, so it is not possible to establish which evidence was considered to provide the rating. No information is afforded regarding the committee's opinion in any decision making on the guidelines, limiting its transparency.

The guideline did not recommend acupuncture, for acute sciatica with an evidence rating of 'A' = a strong recommendation there was no evidence of efficacy for acupuncture. For chronic LBP they do not recommend acupuncture with evidence rating of 'B' = there are doubts as to whether the procedure should always be recommended. However they did recommend manual therapy, manipulation and massage for sub-acute LBP with an evidence rating of 'C' there is deep uncertainty, but reported for acute sciatica it was contraindicated or not useful. For chronic LBP patients' manual therapy was recommended with evidence rating of 'B' (Negrini et al. 2006).

Scotland: Within the UK, the Scottish Intercollegiate Guidelines Network (SIGN) in Scotland develops guidelines and healthcare policies independent to England and Wales and is funded by the NHS Quality Improvement, Scotland. SIGN produced a guideline for chronic pain in December 2013, which encompassed the assessment and treatment of chronic LBP (SIGN, 2013).

The guideline methodology states it should consider if guidelines should be reviewed after a period of three years, though they state having a fixed review time may not be appropriate due to evidence being published at different rates. No reviewed or updated guideline has been published since 2013 limiting the current evidence status of the guidelines.

The guideline development committee developed the guidelines; they consisted of a group of relevant professionals and patient representatives. SR methods were used, and equality impact was assessed, using the AGREE

criteria. The methods were transparent and published as a separate document. The guideline stated its editorial independence from the funding body.

The evidence within the guideline was given a level from 1++ to 4 and was also graded from A to D; A = At least one meta-analysis SR or RCT rated as 1++ (high quality with a very low risk of bias) or a body of evidence consisting principally of studies rated as 1+. D = evidence from non-analytical studies or expert opinion. In addition good practice points were given if recommended best practice was based on the clinical experience of the GDG. Peer review and public consultation was conducted to review the draft of each guideline. The guidelines clearly outlined the evidence used to inform each decision and recommendation.

SIGN used the term manual therapy as an umbrella term for various forms of hands on therapy. They acknowledged extensive research for manual therapy but noted it was generally of poor quality. SIGN recommended manual therapy should be considered for short-term relief of LBP, with evidence graded as 'B' (SIGN, 2013).

SIGN identified systematic reviews incorporating RCTs of varying quality, showing small clinically relevant benefits of acupuncture for LBP compared to waiting list controls. A meta-analysis of acupuncture versus no acupuncture showed overall benefit for LBP. SIGN recommended acupuncture should be considered for short-term relief of LBP with evidence graded as 'A' (SIGN, 2013).

Ireland: No Irish guidelines exist for LBP; it is indicated that Irish healthcare adopts the European clinical practice LBP guideline and their recommendations (Fullen et al. 2007). However, research suggested that though General Practitioners were consistent with prescribing medications, there was less consistency for onward referral than recommended by the European guidelines (Fullen et al. 2007).

United Kingdom: The National Institute of Health and Care excellence (NICE) are a government-funded organisation in the UK whose objective is

to improve outcomes of individuals using the NHS, and other public and social care services, and part of their remit is to produce clinical practice guidelines (NICE, NG59, 2016). The focus of the guideline was on the diagnosis, treatment and management of non-specific LBP in primary and secondary care. The target population was individuals >16 years of age with LBP and/or sciatica of any duration, those whom were pregnant or had consistent LBP following surgery were excluded.

NICE state their guidelines are reviewed for an update every two years with a guaranteed review at least every four years, to ensure their currency and accuracy is maintained in line with the best available evidence. The guideline was reconsidered in 2018 and no new evidence was found to prompt a review or impact its recommendations (NICE, 2018).

The Guideline Development Committee (GDC) was a multidisciplinary team of 12 healthcare professionals and two lay members, who discussed the evidence to develop the draft recommendations of the guideline. SR methods were adopted, and the methods were detailed and transparent, with supporting methodological documentation published separately. The National Clinical Guideline Centre (NCGC) team conducted literature searches (with RCTs given primacy) and prepared systematic reviews and economic analyses. Whenever possible randomised and prognostic data was meta-analysed and reported in GRADE profile tables. Evidence summaries and the GRADE assessments to assess quality of the evidence were produced, for the committee to consider. The committee also invited experts in the field to present at their meetings. The NICE guideline detailed the strength of recommendation within the guideline linked to the strength of evidence presented and reviewed.

However limited information was provided regarding the strength of professional opinion of the guideline development committee or expert opinions, no information is presented detailing how discussions or opinions informed decisions. It is not possible to clarify if recommendations were influenced by professional opinion, or the weight given to the professional opinion comparative to the evidence considered. This potentially allows the

guideline recommendations to be influenced by professional bias of the committee and thus lack transparency.

Prior to publication, a draft version of the guidelines was produced, submitted for review by registered stakeholder groups, and revised if appropriate before the final document was agreed and peer reviewed internally by NICE and ratified. A comprehensive patient and public consultation was not conducted as part of the guideline development process, which was a criticism of the guidelines, however NICE are currently consulting on improving public consultation in the development of their guidance (NICE, 2019). The lack of editorial independence of the NICE LBP guideline remains a limitation, with it scoring 42 out of 100 on the AGREE II assessment conducted for the Canadian guideline (TOP, 2017).

The NICE (2016) guideline recommended manual therapy, including spinal manipulation and massage for LBP. In a change from their NICE 2009 LBP guideline however, they removed their recommendation for acupuncture based on insufficient evidence above sham, and due to not being able to justify the cost of implementation with the limited available evidence. The NICE guideline recommendations are discussed further in section 2.6.

2.5.4 Summary of the clinical practice guidelines

Table 2.9 summarises the key information regarding the scope and methods of the international guidelines as discussed above (section 2.5.4).

Title	Country & Year	Scope of guideline	Methodology	Peer & public review	Committee opinion	Grading of evidence
Clinical Practice Guideline LBP	USA 2017 (2007)	>18 years Acute, sub-acute, chronic, radicular LBP, symptomatic spinal stenosis	Systematic review	Yes Yes	Yes	Yes
HTA guideline LBP	Canada 2017	>18 years LBP of any duration Excluded pregnancy	Systematic review	Yes Yes	Yes	Not directly
European Guideline LBP	Europe 2006	No age demographic specified Chronic LBP	Systematic review	No No	Yes	Yes
KCE Report 287	Belgium 2017	>16 years LBP of any duration	Systematic review	Yes No	Yes	Yes
Dutch National practice Guideline	Netherlands 2013 (2003)	No age demographic specified Non-specific LBP	Systematic review	No No	Yes	No
French Good Practice Guideline LBP	France 2016	No age demographic specified LBP in the working population	Systematic review	Yes No	Yes	Yes
Italian Clinical Guideline	Italy 2006	No age demographic specified LBP of any duration	Systematic review	No No	Yes	Yes
SIGN Chronic Pain Guideline	Scotland 2013	Chronic pain including LBP >12 weeks	Systematic review	Yes Yes	Yes	Yes
NICE LBP Guideline NG59	England & Wales 2016	>16 years LBP of any duration Excluded pregnancy and post-surgical	Systematic review, GRADE, Meta-analysis	Yes Some	Yes	Yes

Table 2.9: Summary of the scope and methodology of the international guidelines

Title	Year	Country	Acupuncture Recommended	Manual Therapy Recommended
Clinical Practice Guideline LBP	2017	USA	Yes	Yes
Clinical Practice Guideline LBP	2007	USA	Yes	Yes
HTA LBP Guideline	2017	Canada	Yes	Yes
European LBP Guideline	2006	Europe	No	Yes
KCE Report 287	2017	Belgium	A recommendation could not be formulated	Yes
KCE Report Vol 48C	2006	Belgium	Yes	Yes
Dutch National practice Guideline	2013	Netherlands	Not considered	Yes
Dutch National practice Guideline	2003	Netherlands	Not considered	Yes
French Good Practice LBP Guideline	2016	France	No	Yes
Italian Clinical LBP Guideline	2006	Italy	No	Yes
SIGN Chronic Pain Guideline	2013	Scotland	Yes	Yes
NICE LBP Guidelines NG59	2016	England & Wales	No	Yes
NICE LBP Guidelines NG88	2009	England & Wales	Yes	Yes

Table 2.10: International recommendations of acupuncture and manual therapy for LBP

The disparity internationally between the recommendation of acupuncture and manual therapy for LBP discussed in section 2.5.4, is summarised in table 2.10. Table 2.10 also highlights if any changes in recommendations between the same guideline occurred over time.

2.5.5 Reasons for the International guideline differences

There are potentially many reasons for the inconsistences that occur across international clinical practice guidelines and these shall be considered further below. Understanding these differences is important to interpreting the clinical practice guidelines and the incorporation of their recommendations into clinical practice.

It could be suggested the variation observed between international guidelines may exist due to the date of publication, lapse of time between the writing of the guidelines and changes driven by the publication of new evidence (O'Connell et al. 2017). However, the Belgium guidelines from 2006 (Nielens et al. 2006), the Canadian guideline from 2017 (TOP, 2017), the Scottish guidelines from 2013 (SIGN, 2013), and USA guidelines from 2017 (Hackethal, 2017) all recommend acupuncture and these four guidelines alone span an eleven-year time period. During a similar time period the European guidelines from 2006 (Airaksinen et al. 2006), and the French guidelines from 2016 (Petit et al. 2016) and the UK NICE guidelines (2016) did not recommend acupuncture. Lack of conclusive evidence of acupuncture beyond sham, was quoted as the reason for not recommending acupuncture, within these three guidelines. The disparity in recommendation from 2006 to 2017, suggests that it is not the time frame and publication of new research evidence that determined any recommendation in the case of acupuncture and manual therapy.

The influence of one international guideline upon another is evident especially with the Belgium 2017 guideline using the NICE 2016 guidelines extensively to inform their guideline. This highlights the importance and need for guidelines to be fully transparent and independent. While

international guidelines consider each other's guidelines, it is evident however they do not always follow the same conclusions.

There are vast differences in quality across the research base for LBP and GDGs are required to generate recommendations when uncertainty of interventions is evident. There is extensive scope for interpretive variation within GDGs and these may be influenced by expertise, personal opinion, public opinion and previous and other guideline recommendations. For example, the American guideline recommended spinal fusion whereas the Canadian and UK NICE guideline did not, and the UK NICE guideline recommended radiofrequency nerve denervation whereas the Canadian and American guidelines established there was insufficient evidence to recommend the intervention (O'Connell et al. 2017; Qaseem, 2017; Juch et al. 2017; Hasenbring, Rusa and Turk, 2012).

The guideline development directive and scope of a guideline may be determinants for the variation, for example effectiveness or efficacy may be a motivating factor. With acupuncture, both the American and Canadian guidelines recommend acupuncture based on comparisons with usual care or no treatment (effectiveness). Whereas NICE (2016) do not recommend acupuncture based on comparisons with clinical effect above sham acupuncture (efficacy) (O'Connell et al. 2017). In addition to this the Dutch guideline (KNGF, 2013) simply do not acknowledge acupuncture within their guideline.

The size of treatment effect threshold expected may also contribute to research being viewed positively or not. In the case of the NICE guidelines 2009 and 2016, the 2009 guideline considered the treatment effect in the effectiveness of acupuncture over usual care not efficacy over sham. Whereas in the 2016 guideline, the minimal clinically important difference (MCID) of treatment effect was increased to 1.0 over sham was introduced (Cummings, 2017), and this changed the threshold of expectation within the research that had previously been included, this is discussed further in section 2.8.

Additional considerations for the differences between the guidelines may include; local traditions, cultural norms, public expectations, local healthcare provision, commissioning, funding, and variations in the guideline development processes (O'Connell, 2017; Dascanio et al. 2015f; Hasenbring et al. 2012).

It may be advantageous internationally for future research to attempt to standardise some of the disparities between the guidelines, to afford more consistency across the guideline recommendations. Pillastrini et al. (2012) for example advocated for future guidelines to adopt the GRADE framework, to aid international comparison and reduce any discrepancies. However, guidelines within countries continue to have differences and are conflicting, therefore gaining some standardisation nationally may aid an initial approach. Such national internal differences will be discussed further in section 2.6, where the differences between the NICE LBP guidelines of 2009 and 2016 are discussed.

2.6 Review of NICE LBP guidelines of 2009 and 2016

This thesis and the pilot study conducted (see chapter four and five) were inspired by the 2009 / NG88 LBP guidelines. The 2016 / NG59 guidelines were published during the writing of this thesis and thus have been included in the below analysis and discussion, to explore the changes and the implications these have on future research planning. The consideration of both these guidelines is key to this thesis as they inform current clinical practice in England and Wales, and will impact upon the requirement of future research into acupuncture and manual therapy for LBP.

Guidelines for clinical practice in England and Wales are developed by the National Institute for Health and Care excellence (NICE). Guideline development committees develop the guidelines by consideration of best evidence surrounding the condition or intervention (Czarnawska-Iliev and Robinson, 2016).

The NICE LBP guidelines are complex documents covering various aspects of LBP management; however they summarise the care pathway with an algorithm shown in figure 2.1.

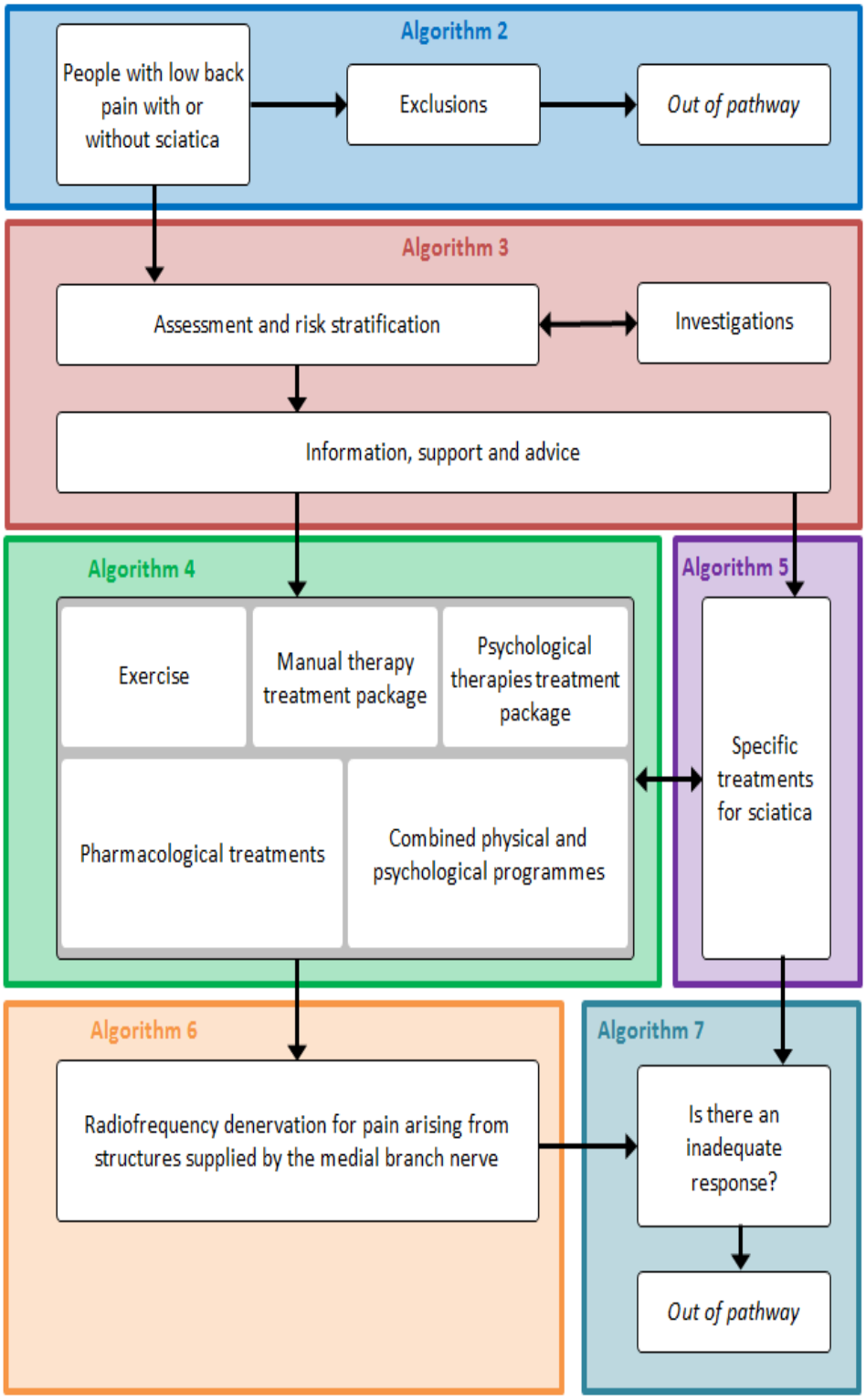

Figure 2.1: Low back pain and sciatica management algorithm (NICE, 2016)

The 2016 NICE guideline is an update of the 2009 guideline, there are some significant differences between the two, as detailed in table 2.11. The more recent 2016 guideline presented the guidelines using a different format to the 2009 and the divergence between the guidelines will be discussed below.

The 2009 guidelines scope covers the assessment and management of non-specific LBP in adults over the age of 18 years, while the 2016 guidelines scope covers the assessment and management of non-specific LBP and sciatica in adults over the age of 16 years, thus extending the target population. No justification for this change is given within the guideline, they simply state:

> *"....they hope to address inconsistent provision and implementation...."*

(NICE, 2016, page 23)

The 2009 guideline provided guidance for the early management of non-specific chronic LBP with a duration of longer than 6 weeks up to one year. Whereas the 2016 guideline includes all LBP from the first episode, (including acute (less than six weeks), sub-acute (six – twelve weeks) and chronic (more than twelve weeks)) with no cut off time point. Their rationale here is to identify the risk of chronicity using the recommended risk stratification tool. This is a novel approach to LBP, moving away from the international norm of timescale classification. The American College of Physicians (ACP) state that while included in their guideline, acute and sub-acute LBP resolves spontaneously for the majority of individuals (Qaseem et al. 2017).

The 2016 guideline also introduced a section on return to work programmes, with the aim of assisting individuals back into work (table 2.11). Key treatment changes have been made regarding medications, most significantly with paracetamol, which is no longer recommended in the 2016 guideline (table 2.11) (Bernstein et al. 2017).

Interventions considered within the NICE LBP guidelines	NICE LBP guidelines (NG88) (2009)	NICE LBP and sciatica guidelines (NG59) (2016)
Risk assessment & risk stratification tools	Not considered in guidelines	Recommended
Imaging	Not routinely recommended	Not routinely recommended
Self-management	Recommended	Recommended
Exercise (individual & group sessions) – including aerobic activity, movement instruction, muscle strengthening, postural control and stretching	Recommended	Recommended
Orthotics & Lumbar supports	Not recommended	Not recommended
Manual Therapy (including spinal mobilisation, spinal manipulation and massage techniques)	Recommended	Recommended
Traction	Not recommended	Not recommended
Acupuncture	Recommended	Not recommended
Electrotherapies (including ultrasound, percutaneous, transcutaneous electrical nerve stimulation, PENS, TENS, laser therapy and interferential)	Not recommended	Not recommended
Psychological therapy	Not included in guidelines	Recommended
Combined physical and psychological treatment programmes	Recommended	Recommended
Return to work programmes	Not included in guidelines	Recommended
Pharmacological therapies	Variations in recommendations: Paracetamol – Recommended NSAIDs – Recommended Weak opioids – Recommended Strong opioids – recommended Tricyclic antidepressants - recommended	Variations in recommendations: Paracetamol – Not recommended NSAIDs – Recommended Weak opioids – Recommended Strong opioids – not recommended Tricyclic antidepressants – not recommended
Injections of therapeutic substances	Not recommended	Not recommended
Radiofrequency denervation	Not recommended	Recommended
Epidurals	Not included in guidelines	Recommended
Surgery	Spinal decompression - Not included in guidelines Spinal fusion – Recommended Disc replacement - Spinal fusion - Not included in guidelines	Spinal decompression – Recommend Spinal fusion – Not recommended Disc replacement – Not recommended

Table 2.11: Interventions considered by NICE for LBP Guidelines

Interestingly in the 2009 NICE guidelines, shown in table 2.11, they do not recommend offering a psychological treatment programme as an individual

intervention for those with LBP, despite some evidence showing them to be efficacious (Gatchel and Rollings, 2008; Turk, Swanson and Tunks, 2008; Chou and Huffman, 2007; Hoffman et al. 2007; van Tulder et al. 2000). They do however recommend the combination of physical treatment with psychological therapy in a combined approach with the aim to address the psychosocial component of LBP for patients. New evidence regarding the psychological element impacting LBP (Pillay, 2016; O'Sullivan et al. 2014; Sadeghian, Hosseinzadeh and Aliyari, 2014; Pincus and McCraken, 2013; Main and George, 2011) has also been published since the 2009 guidelines and psychological therapies as an individual intervention were thus included within the 2016 guidelines. The psychological element of LBP is an extensive and complex area of research and while it is acknowledged, it does fall outside the remit of this thesis.

Another distinct change in the 2016 guidelines was considering and recommending radiofrequency nerve denervation, which was not recommended in 2009. Additionally, acupuncture was recommended in the 2009 guidelines but no longer recommended in the 2016 guideline. The evidence for acupuncture was re-evaluated and the recommendation for acupuncture withdrawn (table 2.11). Of note was that the NICE guidelines for LBP amended their guidance for acupuncture following the consideration of a wider range of research, with some new small-scale studies and more historic research considered and reinterpreted the research with altered parameters, which is discussed further in sections 2.7 and 2.8 of this chapter. Cost delivery of acupuncture was sighted, as one of the main factors but considering an alternative cost-effective delivery model by combining interventions proposed by the pilot study for this thesis was not considered. Poor quality research with questionable results was another of the key reasons for the reinterpretation of the acupuncture evidence. Manual therapy continued to be recommended across both the guidelines.

2.6.1 The influence of NICE guidelines in developing the foundations of this PhD:

The NICE (2016) guideline conclusions, made finding a trial design to address the quality issue for acupuncture studies and a design that eliminated as many potential biases as possible, a more significant need for the future of acupuncture research (see section 2.7.2).

The long-term management of LBP by conventional medical treatments has been shown to be of limited effectiveness (O'Sullivan et al. 2014) and a comprehensive technology assessment and evidence report published in 2010 (Furlan et al. 2010) suggested complementary and alternative therapies should be considered for future interventions, perhaps alongside conventional medicine. The 2010 report considered spinal manual therapy, massage and acupuncture as the three most widely used alternative therapies for LBP internationally (Furlan et al. 2010) and its recommendations were noted to inform the pilot study for this thesis.

The interventions of interest specific to this PhD; manual therapy and acupuncture, were selected in part due to their inclusion within the "Key priorities for implementation" section of the NICE guidance (2009), which included the principle recommendations:

> *"Offer one of the following treatments, taking into account patient preference:*
>
> - *An exercise programme*
> - *A course of manual therapy*
> - *A course of acupuncture*
>
> *Consider offering another of these options if the chosen treatment does not result in satisfactory improvement."*

(NICE, NG88, 2009, page 2)

Though three treatment interventions were recommended within the NG88 2009 guidelines, other than considering patient preference, no indication as to sequential preference of which intervention to offer first was given, nor in which order to offer if multiple interventions were indicated, this was perhaps due to insufficient evidence of superiority of the recommended interventions. The guidelines simply advised that if one intervention did not provide satisfactory improvement, offer another of the treatment options (NICE, 2009).

The lack of clarity and preference within the guidelines potentially leads to a costly approach to NHS healthcare, with patients potentially following several unparalleled treatment pathways. An individual may receive nine manual therapy sessions, followed by ten acupuncture sessions and / or eight sessions of group or personalised exercise programmes (Dascanio, 2015d). No evidence was presented comparing the treatment options to one another, to provide guidance of preferential primary treatment options. No consideration or evidence was given for the potential of combining the recommended treatments options together within one session for a patient. Providing two or three interventions simultaneously within a treatment session would offer considerable cost savings to the NHS compared to offering several separate courses of treatment, as the 2009 NICE guidelines recommend (Dascanio, 2015b). This thesis will aim to investigate this possibility.

Could providing two of the treatments in combination provide an additional benefit to patients and save costs? Could providing one specific therapy first provide preferable results and reduce referral for onward treatment? Is there any benefit of providing additional therapy if one specific intervention has not been successful? All questions, which remain, unanswered by the NICE LBP guidelines (2009) and thus the objective and intention of this PhD was born.

After consideration and planning and with the limited resources of this PhD, the recommendation of an exercise programme was not considered for the pilot study. An exercise programme for LBP falls within part of routinely

delivered care provided by physiotherapists and there are many schools of thought on exercise approaches. Therefore its consideration and investigation falls outside the resources and remit of this PhD. The pilot study focused upon the remaining recommended treatment options of manual therapy and acupuncture and their combination, to investigate the potential for a future full-scale trial.

2.6.2 NICE (2009) definition of acupuncture and manual therapy

> *Acupuncture – is defined as 'the insertion of a solid needle into any part of the human body for disease prevention, therapy or maintenance of health' (Acupuncture Regulation Working Group (ARWG) 2003. NICE, 2009, page 28).*
>
> *Manual Therapy – is defined as a treatment that involves 'manipulation, massage, soft tissue and joint mobilisation.' It is 'a general term for treatments' performed by chiropractors, osteopaths and physiotherapists. (NICE, 2009, page 34).*

2.7 Evidence informing NICE - NG88 2009 and NG59 2016

The following section will consider the supporting evidence across both the 2009 and 2016 NICE guidelines for manual therapy and acupuncture.

The 2016 NICE guideline stated;

> *"The GDG agreed that where interventions have been compared to sham, the sham must be for the intervention of interest e.g. a comparison between acupuncture and sham acupuncture would be accepted however acupuncture compared to sham massage would not."*

(NICE, 2016, page 43)

They went on to say that:

> *"....if there was a lack of placebo or sham-controlled evidence, evidence against usual care will be given priority when decision making."*

(NICE, 2016, page 452)

In addition they stated:

> *"The GDG is required to make decisions based on the best available evidence of both clinical effectiveness and cost-effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is their cost-effectiveness) rather than the total cost alone."*

(NICE, 2016, page 55)

Table 2.12 provides a summary of the research considered for manual therapy and acupuncture for NG 88 (2009) and NG 59 (2016).

	NICE 2009 NG 88	NICE 2016 NG 59
Manual Therapy Classification	Spinal manipulation Spinal mobilisation Massage	Soft Tissue technique Traction Manipulation / Mobilisation Mixed modality manual therapy
Research studies	7 studies -1x high quality 1x moderate quality 5 x low quality – high risk of bias	48 single intervention studies
Research studies continued	Massage 1x systematic review containing 8 papers 1 study – high quality	18 dual intervention studies
Cost Utility economic evaluation	Two evaluations One Manual therapy cost effective One Massage not cost effective Cost and QALY data provided	Five evaluations – all cost effective All with significant limitations Cost and QALY data provided
	NICE 2009 NG 88	NICE 2016 NG 59
Acupuncture	Insertion of solid needle Acupuncture and Dry needling	Insertion of solid needles TCM and Western acupuncture
Research studies	1 x systematic review – high quality 1 x high quality RCT 2 x moderate quality RCTs 1 x low quality RCT	29 RCTs 3 x combined therapy RCTs 2 x Cochrane reviews – not included as did not match protocol criteria
Cost Utility economic evaluation	One evaluation – cost effective Cost and QALY data provided	One evaluation – cost effectiveness Cost and QALY data provided

Table 2.12: Summary of acupuncture and manual therapy research for 2009 and 2016 NICE LBP guidelines

2.7.1 Evidence for manual therapy

NICE guidelines 2009:

Manual therapy interventions are intended to improve mobility, function and pain and are usually delivered by a therapist's hands affecting the

neuromuscular system by the movement of joints and /or soft tissue (NICE 2016; NICE 2009).

Three types of manual therapy techniques were reviewed for the 2009 NICE guidelines and included:

- *Spinal manipulation: A Low amplitude high velocity movement at the limit of joint range, taking it beyond the passive range of movement, (NICE, 2009)*
- *Spinal mobilisation: A joint movement within the normal range of motion, (NICE, 2009)*
- *Massage: Manual manipulation / mobilisation of the soft tissues, (NICE, 2009)*

The 2009 guidelines were informed by evidence from seven RCTs on manipulation and mobilisation techniques, one systematic review and one RCT on massage therapy. NICE (2009) used systematic review methods in accordance with 'The Guideline Manual' (NICE, 2006) to review and assess the quality of the evidence.

SRs had been published on manipulation and mobilisations for LBP, however these were excluded from the 2009 NICE guidelines due to issues of heterogeneity between studies, the appropriate studies were however selected from the systematic reviews and considered individually for the guideline (NICE, 2009).

Of the seven RCTs on manipulation / mobilisation techniques, only one was considered to be of high quality with a low risk of bias (UKBEAM, 2004) and another was considered a well conducted RCT with a low risk of bias (Cherkin et al. 1998). However, five of the trials were considered to be low quality, at high risk of bias. Three of these five low quality studies found no difference between the intervention and control (Hurwitz et al. 2006; Anderson et al. 1999; Doran and Newall, 1975), one found improvement following a two-week follow-up (Triano et al. 1995) and one found significant results but was subject to high attrition (Goldby et al. 2006). The

risk of bias in these five trials centred around the lack of power of the studies, the small sample sizes reducing the precision of any estimated treatment effect, the lack of detailed methodology, the lack of transparency regarding the randomisation process, and the high attrition rate.

Evidence used to support massage included a systematic review of eight international RCTs (Furlan et al. 2002). Though all the studies had small participant numbers, the outcome of the SR suggested a beneficial effect of massage at one year, when compared to sham or self-care education, and an effect equal to the use of corsets. Additionally, one included RCT (Little et al. 2008) judged to be high quality at low risk of bias, found significant improvements between massage and usual care at three months but not at 12 months. While this was a large trial of 579 participants, the participants were randomised to eight separate treatment groups considerably reducing the power of the study; only 75 participants populated the massage group. When considering only the massage group versus the control group (72) this substantially reduced the size of this sub-study and analysis, with the primary investigation considering the Alexander technique.

Although NICE state the evidence suggests manual therapy is more effective overall than usual care, and more effective or the same as other interventions there appeared to be a lack of recent good quality evidence to support the recommendation of manual therapy within the 2009 NICE guidelines.

NICE guidelines 2016:

The types of manual therapy techniques considered by the 2016 NICE guidelines included:

- *Spinal manipulation*
- *Spinal mobilisation (including therapy (SMT) and Maitland technique)*
- *Soft tissue technique*
- *Traction*
- *Mixed modality manual therapy (soft tissue technique +/- traction +/- manipulation / mobilisation*

(NICE, 2016, page 352).

48 studies of single interventions (reported in 55 research papers) and 18 studies (25 research papers) considering the combination of non-invasive interventions with manual therapy as an adjunct were included in the review. This represents a considerable increase in research since 2009, with 22 additional recent studies included, and a greater number of historic papers added which were not included in the 2009 guidelines (NICE, 2016).

NICE (2016) used the GRADE framework to review and assess the quality of the evidence. The outcomes of the studies were summarised individually across the studies, these included pain severity, function, healthcare utilisation, adverse events, psychological distress and Quality of Life (QoL) (NICE, 2016).

Five cost-utility economic evaluations of manual therapy were considered across the range of interventions, all five were considered to have potentially serious limitations but provided cost and QALY data (NICE, 2016).

The recommendations from NICE 2016 were to not offer traction for LBP, but consider offering manual therapy, but only as part of a treatment package including exercise with or without psychological therapy (NICE, 2016, page 452).

The GDG acknowledged there was variable evidence for manual therapy and the wide scope of the intervention as a technique added to the difficulty in assessing it. They reported functional outcomes did not correlate with QoL outcomes. As individual interventions evidence was limited for soft tissue techniques (based mainly on massage), spinal manipulation and mixed modality therapies compared to sham with only short term benefit for some studies (not seen beyond four months). When compared to usual care the evidence was again contradictory and benefits shown were inconsistent across the studies. They did report however a large multi-centred trial reported benefit in QoL and function when manual therapy was combined with self-management and exercise (NICE, 2016).

The GDG concluded the evidence did not allow them to recommend manual therapy as an independent intervention, but it could be considered as part of treatment packages as an optional intervention alongside exercise (Bernstein et al. 2017. NICE, 2016).

2.7.2 Evidence for Acupuncture

NICE guidelines 2009:

The NICE guidelines made their recommendations on the clinical evidence from one systematic review, four RCTs and one cost utility study. NICE (2009) used systematic review methods in accordance with 'The Guideline Manual' (NICE, 2006) to review and assess the quality of the evidence.

The SR was considered of high quality with low risk of bias; it followed SR methods and used the Cochrane Back Review Group (CBRG) guideline (van Tulder et al. 2003) to assess methodological quality of the included studies. It reported acupuncture compared to sham provided some short-term pain relief, and acupuncture compared to no treatment provided some pain relief and functional improvement (Furlan et al. 2005).

One RCT of moderate quality and low risk of bias, reported significant improvement of pain when compared to no treatment, but no significant improvement when compared to minimal acupuncture (Brinkhaus et al. 2006).

The cost utility study found acupuncture to be cost effective (Ratcliffe et al. 2006).

NICE guidelines 2016:

The types of acupuncture intervention considered by the 2016 NICE guidelines included:

- *Traditional chinese acupuncture*
- *Western medical acupuncture*

29 studies (reported in 32 papers) were reviewed; a considerable increase compared to the evidence considered for the 2009 guidelines. These included nine historic papers (published in 1976, 1980 x2, 1986, 1999, 2002 x2, 2003) and eleven more recent (mostly small-scale studies) publications from 2007 to 2014.

As mentioned, NICE (2016) used the GRADE framework to review and assess the quality of the evidence. The outcomes of the acupuncture studies were summarised individually across the studies, these included pain severity, function, healthcare utilisation, adverse events, psychological distress and Quality of Life (QoL) (NICE, 2016).

They reported, one study and two large studies assessed as moderate to high quality evidence, found clinically important benefit of acupuncture. Eight studies showed no clinically significant difference for short or long term pain and a sensitivity analysis of eight studies reported no clinically significant difference between acupuncture and sham. One study showed improvement in healthcare utilisation and another in function (both moderate quality) and a further six studies ranging from low to high quality demonstrated no difference over sham for function (NICE, 2016).

Two high quality studies demonstrated benefit of acupuncture versus usual care for QoL and pain. A sensitivity analysis of eight studies reported clinical benefit of acupuncture but this analysis was rated as very low quality. All evidence comparing acupuncture to active comparisons or as an adjunct to other interventions were assessed as very low quality very small-scale studies (NICE, 2016).

The recommendations from NICE 2016 were to not offer Acupuncture for managing LBP with or without sciatica based on the evidence reported. The GDG discussed the need to demonstrate specific intervention effects, over and above placebo. Though the GDG acknowledged a clinically important improvement in QOL and pain compared to usual care, but as sham evidence is available the sham evidence informs the decision-making and noted no pain relief was achieved beyond four months. The GDG stated that although a

significant number of trials of acupuncture were reviewed the evidence was not consistent in demonstrating a treatment-specific effect. While acupuncture was seen to be superior to usual care, it had not shown significant effect compared to sham, they therefore concluded the improvement was likely to be the result of a contextual or placebo effect. They concluded that extensive evidence ranging from very low quality to high quality evidence was reviewed and that further research was unlikely to alter their conclusions.

2.8 The removal of acupuncture from NICE NG59 (2016)

The removal of acupuncture from the NICE guidelines (2016) occurred following changes in guideline expectations and a re-evaluation of the evidence, with the inclusion of some new (mostly small-scale) studies and historic studies not considered by the 2009 guidelines. For both acupuncture and manual therapy, NICE (2009) considered a small number of studies comparative to the 2016 guidelines. Many of the very low quality and low quality studies were not reported in the 2009 guidelines but were included and considered in the 2016 guidelines.

Cost delivery of acupuncture was sighted as one of the reasons for not recommending acupuncture, due to the high cost to the NHS without evidence of benefit above the context effect. The costs quoted in the guideline for acupuncture were £37 - £56 per hour (NICE, 2016, page 492), at a substantial cost to the NHS of £24,366,000 (NICE, 2009). In respect of this, as mentioned previously one of the considerations I aim to consider with this thesis is the delivery of acupuncture combined with manual therapy in one treatment session, thus a single cost to the NHS. By comparison manual therapy was reported at £51 for an assessment and £39 for a follow-up appointment (NICE, 2016, page 448).

Discussion has occurred regarding acupunctures removal. Some have been pleased as they consider acupuncture as a ‘theatrical placebo” (Colquhoun and Novella, 2013; Cummings, Hróbjartsson, and Ernst, 2018) and others have been less than happy. Of NG59 one such author suggested that:

"...different levels of evidence were used for different interventions, with a perceived bias favouring conventional practice."

(Cummings 2017, page 1)

Cummings (2017) argued that acupuncture was required to demonstrate effectiveness of needling plus context effect versus gentle needling plus context, rather than needling against context of needling. He continued to state that exercise was compared to no exercise, with no context control (usual care) (Cummings, 2017). MacPherson (2017a) supported this view, suggesting the guidelines were compromised by:

"The inconsistent application of the criteria between interventions."

(MacPherson, 2017a, page 248).

NICE reviewed their process for their 2016 guideline to implement consistency with other NICE guidelines (NICE, 2016) and their expectation of a minimal clinically important difference (MCID) of 1.0 on the visual analogue scale (VAS) preferably over sham was applied to their evidence review. This was not conducted for the 2009 guideline.

Some felt however this change of expectation was not universally applied (Macpherson, 2017a; Cummings, 2017; Bovey, 2017; Weeks, 2016; Birch et al. 2016). Cummings, (2017) explained that when compared to usual care, acupuncture achieved 1.61 improvement in pain on the VAS, and when compared to sham, acupuncture achieved a 0.8 improvement in pain on the VAS. Due to the 0.8 improvement being below 1.0, acupuncture received a negative recommendation, despite the improvement of 1.61 above usual care. Whereas comparatively exercise achieved a 0.74 improvement in pain at four months on the VAS above usual care, and exercise received a positive recommendation. (Cummings, 2017; Bovey, 2017; NICE, 2016).

Though statistically these figures are accurate, NICE (2016) debated that sham exercise was not possible to achieve and that remaining active and exercise had been shown to be effective for LBP. NICE reported they were

dubious to support a technique (acupuncture) with inconclusive evidence that may discourage some individuals from remaining active, exercising and caring for themselves (NICE, 2016). This justification did not quiet the critics however and the Acupuncture Now Foundation (ANF) (Weeks, 2016) and others have described the NICE guidelines as 'evidence-biased medicine' due to the discrepancies in their recommendations (Birch et al. 2017; Weeks, 2016).

The comparison of acupuncture to sham acupuncture as oppose to usual care (as for exercise) created another contentious issue in the recommendations (Birch et al. 2016). It was suggested that the GDG did not provide reassurance on their understanding of sham acupuncture not always being inert within some RCTs (Cummings et al. 2018; Macpherson, 2017a: MacPherson et al. 2017c; Cummings, 2016a; Weeks, 2016; Lin et al. 2012; Campbell, 2006). The difference between sham acupuncture and real acupuncture continues to be poorly understood due to the somatosensory tactile component of acupuncture often being included in sham techniques (Makary et al. 2018).

It was argued that due to the physiological mechanism of acupuncture, just moving a needle away from a specific point (non-point needling, often considered sham), or shallow needling or attempting non-penetrating with a Streitberger needle (Strietberger and Kleinhenz, 1998) did not negate the active component of acupuncture treatment and actually has a physiological effect on the body. Sham acupuncture techniques actually have a physiological effect on the tissues they contact due to the innervation of those tissues (Nishiwaki et al. 2018; Chen et al. 2017; Cummings, 2016; Zhu et al. 2013; Lin et al. 2012; Moffet, 2009; Lund and Lundeberg, 2006; Campbell, 2006). Sham acupuncture is discussed further in section 3.5.1.1.

Some studies have shown sham acupuncture to be as effective as true acupuncture (Lowe, 2017; Moffet, 2009), and it has been argued that this is due to acupuncture not being an effective intervention, but merely a good placebo (Cummings et al. 2018; Lowe, 2017; Colquhoun and Novella, 2013). However an alternative argument is that sham acupuncture often maybe an

active treatment (Cummings et al. 2018; Cummings, 2017; Wang, 2017; Lundeberg et al. 2011; Moffet, 2009; Lund, Näslund and Lundeberg, 2009; Lund and Lundeberg, 2006).

Sham acupuncture has been shown to be more effective than usual care, many conventional medications and other interventions (Cummings, Hróbjartsson and Ernst, 2018; Cummings, 2017; Cummings, 2016b; Lundeberg et al. 2011; Moffet, 2009). Until an inert sham for acupuncture can be established, referring to non-TCM and non-point acupuncture techniques as sham acupuncture would be inadvisable for research practice and this is discussed further in section 3.5.1.1.

In addition, it has been suggested that conflicts of interest held by the Guideline Development Committee (GDC) Chair and one other member of the committee were not satisfactorily managed in the public's best interest (Cummings, 2017). It was noted that both members had private practices, which would benefit financially from the recommendation for radiofrequency nerve denervation by the guideline. As previously noted, the technique had previously not been recommended in the 2009 or by other international guidelines based on lack of evidence. Both individuals were new to the committee and took part in the discussion of this intervention and its recommendation (Cummings, 2017; Cummings, 2016). When challenged on the subject, NICE responded advising they felt the declaration of interest made by the individuals was sufficient not to influence the GDGs decision on the matter (Cummings, 2017), even though it received a positive recommendation, despite the inconclusive evidence and other guidelines not supporting the intervention (Juch et al. 2017; O'Connell et al. 2017; Hasenbring et al. 2012).

2.9 Discussion

The literature review I conducted highlighted a lack of high quality evidence investigating the comparison or combination of acupuncture and manual

therapy for LBP. Highlighting the need for further high quality evidence in this area.

It is evident that from the same body of international evidence, a variety of guidelines can interpret the evidence in a variety of ways. Over 39 different international LBP guidelines were reported to exist, and all had inconsistencies to each other across their recommendations, with GDG interpretation of the evidence playing a significant role (O'Connell et a. 2017; Lonnemann et al. 2012). It poses the question are all these guidelines necessary? Would an international collaboration for guidelines be in the best interest of evidence based medicine? How does the research community and GDGs ensure transparency and consistent interpretation of the same clinical evidence?

The disparity across international guidelines in their recommendation of acupuncture is evident, some countries recommend it, other countries do not recommend it and some countries simply do not acknowledge it within their guidelines (Birch et al. 2016). The significant variation in the recommendation of acupuncture may be explained by a variety of factors; the country of origin of the guideline, the type of healthcare practiced / expected, the scope of the guideline, the guideline methods used, how the evidence was appraised, the use of effectiveness or efficacy in the guideline criteria, the minimum threshold of an effect expected above sham / usual care, the research available at the time of writing the guidelines, the influence of conclusions from other published guidelines, the influence of GDC and expert advisers, the transparency of the decisions, and if peer and public review were used. These factors may have a significant influence of the development of guidelines, and the international community should consider how some of these potential differences could be resolved and coordinated to achieve greater international consistency.

Acupuncture is used widely in society, but its use remains controversial (Vickers et al. 2018b). Caution needs to be exercised in assuming acupuncture is effective or not, from reviewing the guidelines, research is currently contradictory, and further high quality research is required.

Establishing the effectiveness of acupuncture in high quality studies and its use in combination with other interventions would add value to the knowledge base.

In the case of acupuncture within the UK, since the 2016 NICE guidelines many physiotherapy NHS departments have seen acupuncture for LBP no longer authorised for practitioners, impacting upon the clinician's clinical autonomy and decision-making abilities for their patients (Davis, 2016; Dascanio, 2015b; Dascanio et al. 2015f). This thesis aims to consider acupuncture and manual therapy independently and in combination for LBP initially through a pilot study, but aims to inform a full-scale study to provide some clarity in the knowledge base and the recommendations in clinical practice guidelines.

The change in recommendation for acupuncture in the NICE 2016 guidelines continues to be a contested decision with the suggestion of a different assessment and expectation approach for different interventions. The effectiveness and use of acupuncture for LBP remains controversial, and the poor quality of acupuncture studies and cost were cited as a reason for its removal and exclusion from the NICE LBP guidelines (NICE, 2016).

Acupuncture and manual therapy are considered as complex interventions and as discussed in chapter one (section 1.6), the design of conducted trials in some cases may not have been sensitive to accommodate their complexity. This may in part explain why many effectiveness studies are inconclusive, present challenges and are of poor quality. Robust trial design is imperative in determining the effectiveness of these interventions.

The need to understand the best trial design to assess acupuncture and improve the quality of acupuncture research, so a decision upon effectiveness can be determined, is apparent. The following chapter will consider the various types of RCT's and aim to recommend a best-fit design to achieve these objectives and thus a trial design for a pilot RCT for this PhD.

3 Randomised Controlled Trials (RCTs)

3.1 Introduction

Thus far I have considered LBP and its impact upon society highlighting the need for further evidence. I have conducted and presented a literature review and identified and discussed systematic reviews of international guidelines for LBP, demonstrating the disparity between the guidelines, their recommendations and lack of conclusive evidence. The NICE guidelines for England and Wales and their update in 2016 have been reviewed and their recommendations for acupuncture and manual therapy discussed.

Chapter two discussed the uncertainty remaining over the use of acupuncture for LBP and the dispute over its removal from the NICE LBP guidelines (2016). Poor quality evidence was sighted as a significant reason for the removal of acupuncture from the NICE LBP guidelines (2016). Many studies in research will fall victim to a random or systematic error (Henderson and Page, 2007). The literature for acupuncture studies has been described as variable, of poor methodological quality, difficultly in interpretation, inconsistent reporting, issues with internal and external validity and high levels of attrition (SAR, 2018; Liu et al. 2015; Qiu et al. 2016; Norvella, 2015; Lee et al. 2012). These issues have often occurred due to poor trial design and the execution of the studies.

The uncertainty that remains surrounding the effectiveness and efficacy of acupuncture is in part due to these issues with study design. Without high quality trials designed for acupuncture, it will not be possible to conclusively analyse it and recommend it for inclusion or exclusion from future clinical practice guidelines and thus the commissioning of current clinical practice.

This chapter will consider RCTs and their intended use, present types of bias and critically review potential RCT designs, in an attempt to understand the strengths and weaknesses of specific designs in relation to bias. It will relate the RCT designs to a potential study of acupuncture and manual therapy

aiming to identify and offer justification for a single design option, which will enable a good quality evaluation and eliminate as many biases as possible to improve future quality studies. Thus potentially offering a solution to rectify the poor quality previously seen in many acupuncture and manual therapy trials as presented by NICE (2016).

The RCT is considered to be the 'Gold Standard' methodology for research evidence (figure 3.1) and randomisation reinforces the validity of the trial. Randomisation reduces selection bias by ensuring any between group differences occur simply due to chance and not due to specific selection of participants (Bondemark and Ruf, 2015; Henderson and Page, 2007). Non-randomised trial designs, by comparison, have inherent sources of bias so limiting internal validity (Kunz and Oxman, 1998).

Figure 3.1: The hierarchy of evidence (Adapted from: http://www.create.ac.uk/blog/2014/07/29/evidence-quality-in-intellectual-property-research-a-comparison-with-the-medical-sciences/)

The essential reason for conducting an RCT is to determine causation. The rigour of the RCT design aims to prevent selection bias, account for regression to the mean, counter temporal changes, and provide a valid basis for statistical inference. However, despite the extensive benefits and

enhanced validity of the RCT, there are potential biases that can be introduced into an RCT, especially if trial design is not rigorous and well executed. All biases need to be attenuated as far as possible to ensure trial internal validity and reduce the risk of a biased treatment estimate (Torgerson and Torgerson, 2008). Koletsi et al. (2011) reported in their study of RCTs that while 112 studies identified themselves as RCT's only 33 (29.5%) could be classified as RCTs through their demonstration of adequate random number generation and allocation concealment, concluding there is a clear need for the improvement of accurate reporting of RCTs.

RCTs have also been shown to be onerous in their planning, funding and execution, with recruitment often being a significant problem encountered. In a review of 114 RCTs, delay in the commencement date occurred in 41% of trials and recruitment problems were experienced in 63% of trials. Less than 31% fulfilled their targeted recruitment reducing the power of the trials and 53% of the trials used a recruitment extension (McDonald et al. 2006).

3.1.1 Explanatory and pragmatic trials

RCTs can be an explanatory or pragmatic in their design, with each design addressing differing research questions (White et al. 2002a).

Explanatory trial: An explanatory trial aims to explain the specific effects and mechanism of an intervention under 'ideal' conditions (Thorpe et al. 2009), tightly controlling for all non-specific elements. These trials compare an intervention to a placebo control and are often used in drugs trials and the early development of interventions. Explanatory trials have high internal validity however lack external validity and the ability to generalise to clinical practice (Sedgwick, 2014a; Greenhalgh, Howick and Maskrey, 2014; MacPherson, 2004).

Pragmatic trial: Pragmatic trials aim to establish the effectiveness, risks and costs of an intervention within its routine environment, without separating out the specific and non-specific effects of the treatment (Thorpe et al. 2009). These trials have a greater population inclusion, with fewer inclusion criteria

and exclusion criteria, they have a wider potential range of study sites, they are less tightly controlled, allowing a wider view of an intervention and they utilise approaches to improve effectiveness as would be the case in clinical practice. These attributes afford greater external validity, more closely reflecting and informing intervention use within clinical practice, however this can be at the expense of the internal validity compared to explanatory trials, (Ford and Norrie, 2016).

3.2 Potential bias in an RCT

Bias within trials can alter our estimate of the true effect of the investigated intervention. In scientific research, bias implies the observation of an effect that occurs due to a systematic error. An example would be if an individual were encouraged or selected (in a recognised or unrecognised way) one of the investigated outcomes over the others due to influence on their judgement (Pannucci and Wilkins, 2010). The effect of bias can occur in both directions to overestimate or underestimate the true intervention effect (Cochrane, 2018), it is important to understand the risk of bias and eradicate it as far as possible within RCTS.

In the following section the potential biases that occur within an RCT will be discussed and how these impact upon the research process. While other biases do exist for example commercial bias (Ahn et al. 2017), and there is extensive evidence to suggest that trials and other studies can be very biased, the following section will focus upon the bias that can be controlled for, in a study for this PhD.

3.3 Sources of bias in RCTs

3.3.1 Selection bias

Selection bias occurs when baseline characteristics are systematically different between intervention arms. Differences that occur within study

arms need to occur due to chance not due to an outside intervention. Adequate allocation concealment and effective randomisation can reduce this bias (Cochrane, 2018; Henderson and Page, 2007).

3.3.2 Performance bias

Performance bias is the systematic unintended differences occurring between intervention arms, often due to lack of blinding. An example would be if a therapist knew they were delivering the true intervention they maybe more positive in their delivery and communication than if they knew they were delivering a placebo treatment. Equally if a participant thought they were receiving the true intervention they may be more encouraged than if they believed they were receiving a placebo (McCambridge et al. 2014; Rubinstein et al. 2014). Adequate blinding in placebo trials or equivalence of interventions in comparison trials can reduce this bias.

3.3.3 Detection bias

Detection bias refers to the process of outcome assessment and how these may differ if not standardised or outcome assessors not blinded. Detection bias does tend to be less of an issue for pragmatic RCTs as standardised outcome measure tools are used, however this does not always eradicate the problem and blinding outcome assessors can be an effective solution (Viswanathan et al. 2013).

3.3.4 Attrition bias

Attrition is the loss of participants during a trial that cannot be incorporated into the trial analysis. Attrition of between 0 - 5% is not considered to substantially affect the trial results to any degree of significance. However, attrition levels are purported to average approximately 20% and are generally poorly reported or addressed (Dumville, Torgerson and Hewitt, 2006). Though some differential attrition is likely to occur within all trials, non-

random attrition across trial arms is the key concern. The non-differential attrition within trials can significantly reduce the validity of trials (Strite, 2010).

Various reviews have reported a range of attrition rates ranging from 4 – 28% (Hewitt et al. 2010) and 7 – 67% of attrition within trials (Dumville et al. 2006; Leon et al. 2006); these substantial rates are likely to influence the outcome of their trial. A study of CBT for LBP reported an attrition rate of 18% (Glombiewski, Hartwich-Tersek and Rief, 2010); a trial of manipulation and exercise for LBP reported 3 – 6% attrition (Evans et al. 2018). A study of acupuncture and sham acupuncture for LBP reported a 4% attrition rate, however a loss of 10 and 11 participants from the acupuncture and sham acupuncture arms and a loss of 25 participants from the conventional therapy arm was seen (Haake et al. 2007). While some reported attrition figures are low, if disproportionate attrition occurs across the intervention groups it may introduce post-randomisation selection bias, potentially altering the direction of the treatment effect, leading to a misinterpretation of the trial results.

If random attrition occurs and the loss to follow up rate is similar between intervention arms, any potential attrition bias is minimised (though the reduction in numbers will reduce the statistical power of the study). Further attrition bias can be reduced by thorough efforts to follow up all participants and retain participants even if they do not elect to continue with their randomised intervention. The utilisation of an intention to treat analysis would aim limit any bias resulting from attrition and selection bias (Torgerson and Torgerson, 2008).

The majority of attrition often occurs within the early stages of a trial, usually between the collection of baseline data and the first follow-up point (Torgerson and Torgerson, 2008). The MRC/BHF (2002) trial of antioxidant vitamin supplements for the prevention of cardiovascular disease reported almost one third of the total participants dropped out within the first time period (a run-in period) of their trial. A trial investigating McKenzie physiotherapy for low back and neck pain, with a 12 month follow up period, reported 75% of the attrition occurring within the first follow up period (6

weeks), with 18% of participants being lost. The accumulative attrition at 6 and 12 months increased from the reported 18% to 23% and 24% respectively but a further increase of only 5% and 1% attrition (Klaber-Moffett et al. 2006). Three further trials on back pain reporting attrition, investigated spinal manipulation (UKBEAM, 2004), cognitive behavioural therapy (Lamb et al. 2010) and Yoga (Tilbrook et al. 2011) and presented similar rates of attrition at 25%, 22% and 13% respectively.

Scope for potentially limiting and reducing attrition in the early phase of a trial may be achieved by the use of a run-in period or a trial nested within a cohort study; these designs are discussed in section 3.4.5 (run-in trials) and section 3.4.7 (cohort design with nested RCT).

3.3.5 Reporting bias

Reporting bias signifies the differences between reported and unreported results. The reporting occurs within studies due to the investigators reporting the significant differences in publications, but less likely to report differences, which are not significant (Cochrane, 2018). Chan (2005) suggested this was the most common form of bias within RCTs. Publication of all trial results both positive and negative should be encouraged and supported to reduce this bias (Mlinaric et al. 2017; Goodchild, 2015).

3.3.6 Resentful demoralisation

Participant preferences can introduce an outcome bias. It is thought that an individual's treatment preference may alter their outcome and therefore the trial results (Bowling and Ebrahim, 2008). This can occur in two ways, a participant may be very encouraged by their randomised intervention and approach the trial with a positive attitude to the intervention they wanted, as the key motivator for entering a trial may be to receive the trial intervention. This overtly positive approach may influence the treatment effect, thus potentially more positively skewing their results. Alternatively, if not allocated to an intervention they wished for, participants may be despondent

and exhibit resentful demoralisation bias (Brewin and Bradley, 1989), or even drop out of the trial contributing to attrition bias of the results.

UKBEAM (2004) reported compliance to the manipulation intervention at 92% but a lower rate at 63% for the exercise intervention. Potentially indicating participants entering this trial had a preference for the manipulation intervention, or that the intervention may potentially be easier to adhere to. Another trial of active treatment interventions for neck pain demonstrated participant preference did affect their outcome (Klaber-Moffett et al. 2005). However, a trial of exercise for LBP offering participant preference for exercise versus usual GP care indicated no difference in outcome of participants who preferred the exercise intervention with those who were indifferent to their randomised intervention (Klaber-Moffett et al. 1999). A trend of reduced compliance or indifference to an exercise intervention may be an area of further exploration, as if resentful demoralisation does have greater impact on exercise comparative to other active interventions, having an exercise comparator in effectiveness trials could potentially lead to an inaccurate reporting of the treatment effect within those trials.

A systematic review conducted by a research collaborative in 2008 reported participants assigned to their preferred intervention having a greater effect size than those who were indifferent or who received their undesired treatment. It concluded prior to randomisation, participant preferences should be undertaken (PCRG, 2008). Offering participant preference within a trial is one potential method to reduce resentful demoralisation and other bias. Participant preference trials are discussed further in section 3.4.4.

3.3.7 Outcome ascertainment

Outcome ascertainment is triggered when investigators over report the effects of one intervention of a trial over the other, this can occur consciously or unconsciously. If this occurs bias will be introduced into the results and could lead to the over or under inflation of a reported treatment effect.

A systematic review of 21 trials analysed ascertainment bias of blinded and non-blinded outcome assessors on their assessment of treatment effects. It reported an over-estimate of odds ratios by an average of 36% by non-blinded assessors of subjective binary outcomes, demonstrating substantial bias in the estimation of effects of non-blinded assessment (Hróbjartsson et al. 2012).

Ascertainment bias can be limited by the blinding of assessors or by collecting the results from a number of sources, i.e. having two doctors assessing a blinded condition post intervention or alternatively the use of patient reported outcome measures independent of the practitioner and researcher (MacPherson et al. 2008).

3.3.8 Hawthorne effects

The Hawthorne effect is the improvement a participant gains as a direct result of being part of a trial or experiment (Sedgwick, 2012b). This was recognised in an experiment on factory workers where productivity improved, whether working conditions were either improved or worsened, improvements in productivity resulted due to being part of an experiment and the influence monitoring (Silverman, 1998).

The Hawthorne effect can be minimised by using a sham or placebo intervention, or alternatively through the comparison of the intervention to an alternative intervention in an effectiveness trial. The objective being to ensure comparability across groups, thus the Hawthorne effects would be anticipated to be the same for all and any effects counterbalanced (Torgerson and Torgerson, 2008).

3.4 Trial Designs

In the following section I consider and discuss potential RCT designs, with the aim to identify the advantages and disadvantages of each trial design with reference to minimising the biases described above. The trial designs

considered; placebo, crossover, factorial, and preference RCTs, RCTs with run-in periods, adaptive cohorts and cohorts with nested RCTs.

3.4.1 Placebo randomised controlled trials

Placebo randomised controlled trials are generally respected as the 'gold standard' of RCT design in health care research. They are considered as the most reliable method for evidence-based medicine and play an invaluable role in the development of healthcare (Spieth et al. 2016; Kaptchuk, 2000). They are increasingly becoming an expectation of guideline developers, such as The National Institute for Clinical Excellence (NICE) in the UK and also by national regulators such as the Advertising Standards Authority (ASA).

In addition to the advantages attributed to all well designed RCTs, placebo RCTs are considered a preferential design due to their ability to reduce or eliminate potential post randomisation biases including: ascertainment bias (through blinding); Hawthorne effect (all patients receiving the same level of care); resentful demoralisation (as patients are blind to their allocation) and potentially attrition bias (as individuals are unaware of their allocation so any attrition should be random and theoretically evenly spread across trial arms). Consequently, their robust nature of limiting bias increases the validity of the results, making them the trial design of choice and the reason many in healthcare assign credibility only to RCTs with placebo design (Torgerson and Torgerson, 2008).

The strength of placebo RCT partly lies in its design eradicating the 'placebo effect' in the assessment of a drug or intervention. Effectively there is blinding of trial clinicians and participants, affording increased the internal validity and reliability of the trial results (Relton et al. 2010).

The history of placebos dates back to the 18th century (Jütte, 2013), and it is reported that placebos, due to limited knowledge of disease aetiology, may have been significantly improved the health of many (Munnangi and Angus, 2019). Placebos have been used extensively throughout research studies and are reported to work on mechanisms modulated by the brain, such as pain,

and are stated to make participants feel better but not cure them (Kaptchuk and Miller, 2018). A placebo or sham intervention is a 'fake' medical treatment or procedure designed to convince a participant in a clinical experiment they have received the real treatment. A placebo is proposed to cause a psychological improvement rather than a physiological one and the placebo effect occurs when a participant's condition improves (or deteriorates) in response to a placebo treatment (Munnangi and Angus, 2019). Though some in research contests this theory, they believe the psychological can affect the physiological and this is discussed further below.

Blease (2012) argues that the placebo effect should be re-termed the "positive care effect", as all interventions occur in the 'context of care' and this can have a powerful effect on an individual. The 'context of care' incorporates all the aspects associated with an intervention but not the intervention itself, i.e. the method of delivery, the location of the intervention, the interpersonal relationships between the patient and the therapist, the patient's belief in the treatment, the therapist's belief and communication regarding the intervention (Blease, 2012). The expectations of a patient are known to play a significant role in the success of a placebo effect (Benedetti, 2013).

The strength and effect of a placebo has been shown to vary for individuals and conditions, with four areas reported to be involved in a placebo effect, 1) The expectation and conditioning, 2) The placebo effect on the brain, 3) Psychoneuroimmunology (the effect of brain activity on the immune system), 4) Evolved health regulation (the body's ability to respond to illness). Placebos have been shown to be effective in many situations and have also been demonstrated to improve the effect of current treatments (Newman, 2017). Beecher (1995) used the term 'The Powerful Placebo" due to its positive effect in many clinical trials.

In the medical field if an intervention emerges as equivalent to a placebo treatment it attracts dismissal or refutation by the research community, however this view may be changing. Scientific research has shown that in addition to the modification of subjective (psychological) symptoms, placebos can elicit specific relevant regions of the brain to be stimulated and

various neurotransmitters (e.g. endorphins, dopamine, cannabinoids) are released, giving placebos a valid legitimacy (Kaptchuk and Miller, 2018). Research has shown that placebo effects account for more than unexplained recovery or regression to the mean (Katchup et al. 2008). Some researchers are starting to investigate how to harness the placebo effect for interventions, rather than dismissing or trying to minimise its effects (Newman, 2017). Critical questions are therefore raised; should placebo treatments be embraced, can placebo treatments be used legitimately in some cases, rather than potentially harmful medicines.

Despite their many advantages however, placebo-controlled trials have been criticised. Justification for using placebos when valid treatments are readily available is ethically questionable and concerns regarding participants understanding of placebos have been raised. Although they offer high internal validity, they are often explanatory trials in nature and may have limited external validity. Their controlled nature reduces the ability to emulate 'normal' environments and thus limits their generalisability to clinical practice (Blease, Bishop and Kaptchuk, 2017; Frieden, 2017; Ford and Norrie, 2016; Spieth et al. 2016; Keränen et al. 2015; Greenhalgh et al. 2014; MacPherson and Thomas, 2007).

Enrolling a participant into a placebo trial may not reflect usual clinical practice where both the patient and the clinician are fully informed of the treatment being offered, limiting the clinical application of a placebo-controlled trial (Blease et al. 2017; Ricker et al. 2015). Clinicians may also consciously or unconsciously promote the 'placebo' effect within everyday practice as an additive effect to complement and enhance the psychological and physiological effects of treatment. Tightly controlled placebo RCTs do not allow this natural enhancement to occur and the effectiveness of a therapy as it would occur in clinic may be underestimated (Bystad, Bystad and Wynn, 2015). Additionally there is significant heterogeneity in the placebo interventions used in trials, limiting standardisation and analysis across trials. The effect sizes of placebos and sham techniques have been shown to vary with some potentially having active treatment components (MacPherson et

al. 2017c) (discussed further in section 3.5.1), and inconsistencies in placebo effects have been demonstrated in some NICE network analyses (White and Cummings, 2012).

Placebo RCTs are explanatory trials in their nature, and establishing effectiveness of a clinical intervention may well be served better with an alternative design. Pragmatic trials have been suggested as potential alternatives to placebo RCTs, they can limit bias while still retaining the characteristics of clinical practice, allowing the wider application and external validity of any findings (Ford and Norrie, 2016; Relton et al. 2010).

While there are arguments for and against the use of placebo RCTs, it is important to recognised that in research or clinical practice, all real or sham interventions are delivered in the context of care with the routine of treatment, and these context effects and routines need to be considered in research (Kamper and Williams, 2013). The full theory and study of placebo is a vast subject area and as such beyond the realms of this PhD, an introduction and key principles of placebo are considered here and further discussion of placebo RCTs is presented in section 3.5.1.

3.4.2 Randomised crossover trials

A trial design widely used by pharmaceutical companies is the crossover design (Mills et al. 2009). Participants act as their own control in crossover RCTs, they receive both interventions but in a random order. Commencing in one group, then after a specified time period to allow for washout, crossing over to the other intervention. This may consist of an intervention and a control, or two or more interventions. A benefit of this design is that a paired data comparison can be made (participants compared with themselves), reducing the variability and thus increasing power. Fewer participants are required, as they act as their own controls. These can only be used for chronic conditions however where treatment is for symptom relief, not cure (Senn, 1993).

Crossover trials have been shown to be useful for chronic pain trials of medication. Pain relieving drugs have a wash out period therefore after a short delay appropriate to the carry over effect; the implementation of the alternative treatment can be administered (figure 3.2) (Sibbald and Roberts, 1998). With all participants receiving the same treatment regime just in a varied order an added advantage is establishing if treatment response is influenced by participant characteristics.

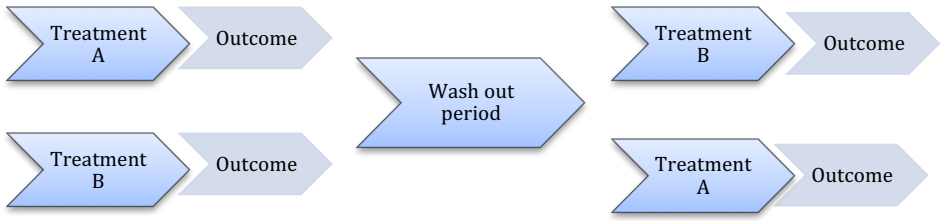

Figure 3.2: Design of a crossover trial *(adapted from Bowling and Ebrahim, 2005)*

Crossover trials are recommended for stable and chronic conditions, such as arthritis, their use is usually for interventions with short lasting effects (Mills et al. 2009). One issue with cross over trials is the potential 'carry over' (and differential carryover) effect of an intervention or the lack of wash out of the intervention. The wash out periods can increase trial length and create ethical concerns for any trial participants who potentially are without treatment for a period of time (Yang and Stufken, 2008).

In trials of medicinal products physiological measures via blood testing can help ensure the washout effect has occurred. This is more difficult to determine in non-pharmaceutical interventions, it is not known what the wash out effect of active interventions are, and no test exists to measure clinical effects. MacPherson et al (2017b) reported the effect from acupuncture trials remaining at 12 months. It would thus be difficult to establish if any treatment effects had been completely eradicated before the commencement of an alternative active treatment intervention. Crossover trials would not therefore

be recommended for the comparison of complex active interventions with an unknown period of the washout effect.

3.4.3 Factorial RCT design

A factorial RCT design allows for two or more interventions to be analysed against a control separately and also in combination (table 3.1). It also allows for the potential assessment of an interaction between interventions (Pocock, 1983). An advantage of this design is the reduced cost; the factorial design will enhance the cost effectiveness and efficiency of a trial by allowing the analysis of two interventions concurrently while still including a control arm, effectively delivering two trials for the relative cost of one (Finlay et al. 2003).

Table 3.1 demonstrates the factorial design, with the central four cells indicating the treatment groups for allocation. The margins highlight how the comparison analysis of the results would occur.

Interaction between treatments and over usual care can also be detected with this design, analysing for synergy (or additive effects), or negative effects of a combination of treatments. This makes it an effective method for comparing complex interventions while still maintaining a control arm.

	Treatment B	**No Treatment O**	**Margin**
Treatment A	AB	AO	All A - AB & AO
No Treatment O	BO	OO	All non-A - BO & OO
Margin	All B - AB and BO	All non-B - AO and OO	

Table 3.1: Factorial 2 x 2 trial design, with analysis at the margins

It has been suggested that this design is under used and should be adopted more widely (Torgerson and Torgerson, 2008) however possibly due to its more complex design and analysis, researchers have not embraced it. It has

also been shown that many factorial RCTs are often underpowered for the interaction analysis, limiting their validity (McAlistair et al. 2003).

3.4.4 Preference trials

Preference trials are designed to offer participants a choice of interventions, providing the preferred intervention to those who have a preference and randomising any participant who has no preference (Brewin and Bradley, 1989). Offering a preference aims to decrease resentful demoralisation and maintain participation by providing the participants their choice of treatment with the intention of minimising any crossover effects to usual care and dilution bias (Bowling and Ebrahim, 2008).

Preference trials however, do require a greater sample size and have an increase cost. A substantial number of the participant population can receive their preferred intervention but be lost for analysis from these studies, because if they chose their preferred treatment and had a strong preference, they are not able to be randomised, only participants with no preference are randomised in these designs, potentially also reducing the external validity of the trials (Torgerson and Torgerson, 2008).

However Bradley-Gilbride and Bradley (2010) consider that offering preference could increase the external validity of an RCT. They suggest the preference design increases inclusivity of the population, with a greater range of individuals agreeing to enter a trial because those with a strong preference who would often refuse to enter an RCT have less fear of receiving unwanted interventions would agree to enter.

Preference participants are often not included in the analysis of the trial interventions as they have psychological expectations of the intervention, which may introduce bias, including selection bias. One solution to this is to determine individual preference but continue to randomise all participants, this permits analysis of whether an original preference influences outcome. An example of this design was the SPRINTER trial (2005) (figure 3.3).

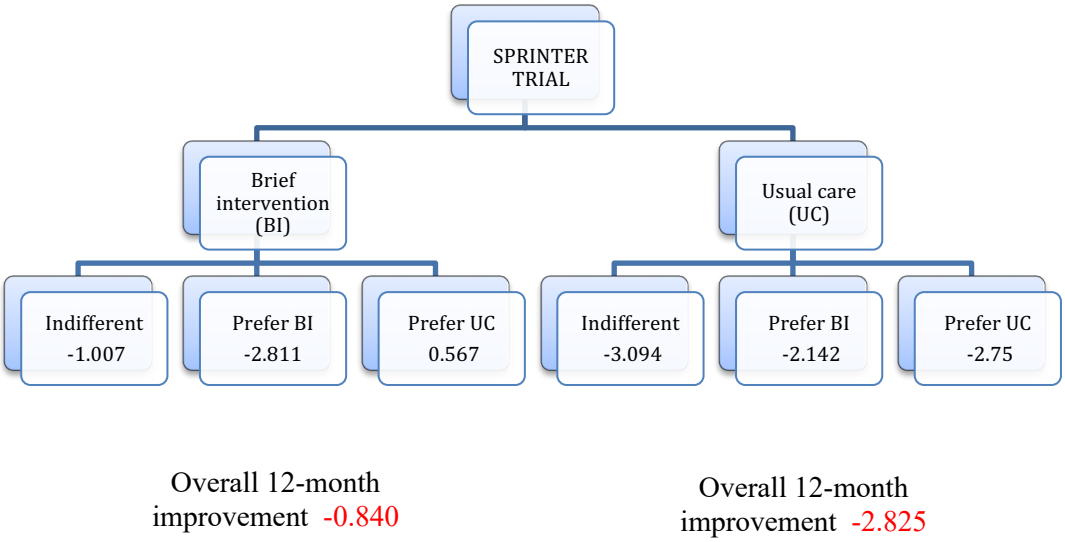

Figure 3.3: Flow diagram of SPRINTER preference trial results

(Modified from www-users.york.ac.uk/~djt6/lecture-preference-trial.ppt)

The SPRINTER trial results demonstrated that those who preferred the brief intervention had a greater clinical improvement than those who preferred usual care. Participant preference to an intervention was shown to impact upon the trial results, with analysis demonstrating interaction (SPRINTER, 2005).

Though all participants being included in the trial analysis increases the external validity of this design, resentful demoralisation and attrition may be increased reducing the internal validity. As individuals are asked for their preference then randomised, and thus not necessarily given their preferred treatment choice, being asked for their preference and then potentially given something else could potentially increase resentful demoralisation further (Torgerson, Kaber-Moffet and Russell, 1996).

Zelen's method or Pre-randomised consent:

Zelen's method and pre-randomised consent are types of preference trials. Their approach potentially deals with post-randomisation biases as they specifically focused upon participant preference. Zelen's method is a design in which participants only consent to treatment after randomisation has occurred. However, this poses difficulty for some in the research community

who feel this is an unethical approach (Relton et al. 2010; Campbell et al. 2005; Stevens et al. 2001).

Participants, who consent to treatment post randomisation and have been allocated to the intervention, receive the intervention, those who do not consent to treatment even, if allocated, receive standard care. The intention to treat (ITT) principle should be used in Zelen's design, so all participants are followed up in the study in the intervention group they were allocated to, regardless of the intervention they received, consented to or crossed over to, thus avoiding selection bias (Yelland et al. 2015; Gupta, 2011; Adamson et al. 2006).

Consenting post randomisation in this design does risk participants refusing their allocated treatment and thus may introduce a dilution of the treatment effect following the ITT principle. However, ITT is not always used in designs of this nature (which itself would introduce selection bias) and a review of trials using Zelen's methodology showed 29% did not follow the ITT principle, and thus reported potentially biased results (Adamson et al, 2006).

Zelen's design also aims to prevent resentful demoralisation and subsequently reduce attrition, with those participants randomised to the control group / standard care not being advised of their allocation, that they are in a control group or even part of a trial. This provides an alternative to a patient preference trial design (Torgerson and Torgerson, 2008).

Disadvantages of this design are that participants may cross over. If a participant is randomised to receive an active treatment, they may refuse the treatment when it is offered, therefore crossing over into the usual care or control group. This was reported to occur for 10 - 36% of participants, a significant percentage of crossover (Altman et al. 1995). This can lead to crossover attrition and following the intention to treat analysis model would impact on treatment findings, causing dilution bias and reducing the power of the study.

Additionally, ethics committees are not supportive of this design, considering it unethical not to inform the participant which treatment group they have been allocated to or that they are even part of a trial (Torgerson, 2001). For these reasons Zelen's method is rarely used (Adamson et al. 2006).

3.4.5 Run-in period, delayed-start or pre-randomisation time period trials

Run-in periods, delayed start or pre-randomisation time periods can be incorporated into trial design. They provide time for evaluating participant's symptoms or disease and thus evaluate the interventions actions upon those. Thus the time element of the study can help to distinguish whether a reduction of the symptoms or modifying of a diseases progress is significant. The design has been used successfully in studies of Alzheimer's or Parkinson's (D'Agostino, 2009). The run-in trial is another trial design that can aid the reduction of attrition within RCTs. Participants are recruited to a trial but not randomised until after their first follow-up period has been completed. Previous trials have shown that the majority of attrition occurs within the first follow-up phase and therefore randomising participants after this time point, should reduce attrition rates and selection bias (Torgerson and Torgerson, 2008).

The MRC/BHF study group (2002) used this design when considering vitamin supplements for high-risk individuals with heart disease. They reported during their run-in period almost one third of participants dropped out of the trial. Randomisation after the run-in period reduced their attrition and potential bias, thus improving the validity of the outcome of the trial.

It could be argued that this approach may lead to the selection of only the more willing participants and hence provide an overestimation of the population treatment effects of a trial, reducing the external validity. However participants otherwise included would be lost prior to randomisation, and those lost may be of similar persona, potentially reducing the diversity of the trial population. The participants lost are likely those who would have likely dropped out of the trial after randomisation and their results

would then follow the ITT principle, potentially diluting the overall treatment effect and outcome. Though trying to retain participants and include a diverse population is important, the reduction of attrition would arguably be most beneficial in trials (Dascanio, 2015c).

A disadvantage of a design with a run in period or delayed start is that some participants may lose interest while waiting for an extended period following their acceptance into the trial, thus attrition during the run-in period may be considerably higher, comparative to studies that randomise to the intervention at the acceptance into the trial (Velengtas, Mohr and Messner, 2012).

An additional concern is that the use of run-in periods may lead to an overestimate of the impact of the treatment on the population, as in a trial situation some participants would, when offered the intervention, refuse or drop out. However, in a clinical setting, the clinician could be recommending the intervention to the patient and they may be more likely to follow a clinician's advice. Also, if a treatment were offered on the NHS on the basis of a trial using a run-in period, the results are based on the participants waiting for run-in period before receiving the intervention and for a similar accurate effect the patients may also need to wait for the set period prior to receiving their recommended intervention to reproduce the trial outcomes. NHS waiting times however may render this less of a problem.

Trials with run-in periods and those without may produce conflicting results. If significantly different levels of attrition occurred between trials, greater attrition and following ITT principle would inevitably lead to a dilution bias of the trial results, thus providing conflicting estimates of effect.

3.4.6 Adaptive cohort design

An adaptive cohort design is a study that facilitates modifications to occur to the interventions and / or statistical process during the study period, while maintaining the reliability and validity of the study (Chow and Chang, 2008). They are 'standard' RCTs but with built in processes to allow planned changes to occur. They were designed primarily for the pharmaceutical

industry in response to cost and time concerns for the evaluation of medicines. They aid the examination of several medicines and allow the implementation and withdraw of medications as the results during the trial indicate, thus providing vast cost saving on trials and accelerating the evaluation process (Mahajan and Gupta, 2010). Maintaining a sufficient cohort size and statistical power within these studies can prove challenging but prior study data from previous phases can be used during statistical analysis (Lestini, Dumont and Mentré, 2015).

Adaptive cohort designs can be very useful if one is testing several promising interventions and the design enables the trial to be finished early in the event of futility, inferiority or superiority. They have an advantage of flexibility of pre-planned modifications (set at protocol stage and approved by ethics) that can be applied to the trial, for example; increasing the dose of a treatment, adapting the treatment duration, providing additional input (Lestini et al. 2015).

The adaptive cohort design may benefit a participant with an aggressive disease, where clinicians are looking for immediate responses to therapies. Additionally, if an intervention is deemed to be ineffective or detrimental then the intervention arm can be ceased early (Mahajan and Guptal, 2010).

Adaptive cohort designs have included the introduction or withdrawal of interventions during the course of a trial. This was used in the STAMPEDE trial (James et al. 2008), a cohort version of the adaptive RCT design. STAMPEDE was developed for the treatment of prostate cancer, comprising of six arms and five phases, this design methodology was used to answer research questions as quickly and reliably as possible (James et al. 2008). When one form of medication was deemed ineffective during the trial, it was withdrawn and not offered to any further recruited participants.

With some adaptive designs if a new medication becomes available during the trial it can be added to the trial and incorporated into the randomisation process for new participants. The nature of this design allows it to be very

flexible and adapt to the current and new treatments available for conditions (Sydes et al. 2009).

There are however significant challenges in the practicality of the adaptive cohort design and the statistical analysis of such a complex design, due in part, to the changing variable parameters and any early stopping. The multiple layers of the design increase the studies complexity and the results can become difficult to interpret (Chow and Chang, 2008). Sydes et al. (2009) reported that effective planning and statistical expertise can overcome the challenges to ensure the design provides effective evaluation, they argue these designs though complex should be more widely adopted.

Disadvantages of the adaptive cohort design include the possible overestimation of the benefit (or harm) if a trial is ceased early due to superiority or inferiority, as regression to the mean may have exaggerated any effects (Bassler et al. 2007). The statistical analysis involved in these trials is very complex and requires expertise to ensure accuracy. Multiple interim analyses may also jeopardise the integrity of the study (Mahajan and Guptal, 2010; Sydes et al. 2009). Also due to the nature of the design and over recruitment, if the trial runs to fruition, the increased length of recruitment and sample size is usually larger than an average trial and runs for longer thus making it potentially costlier than a conventional RCT (Lansberg et al. 2016).

3.4.7 Cohort design with nested RCT

The cohort design with nested RCT combines the strengths of a randomised trial with an observational epidemiology study (Loannidis and Adami, 2008) as shown in figure 3.4.

This design has the potential to address issues of recruitment, attrition, resentful demoralisation, patient preference, ethical challenges, eliminating the need for advising participants they are in the control group and allowing treatment comparisons (Bibby et al. 2018; Relton et al. 2010), they may therefore provide a solution to many of the issues highlighted in the other designs.

In this design, only compliant participants are randomised into the RCTs thus reducing attrition. As previously discussed, attrition is commonplace in the early phases of trials; in this design the RCT is nested within the cohort and participants will have completed at least one follow-up questionnaire prior to the RCT commencing. Therefore, any non-compliant participants are likely to drop out prior to randomisation. Reduction of attrition reduces the sample size requirement and there is also less need for the over inflation of sample size, often used to account for any attrition (Relton et al. 2010).

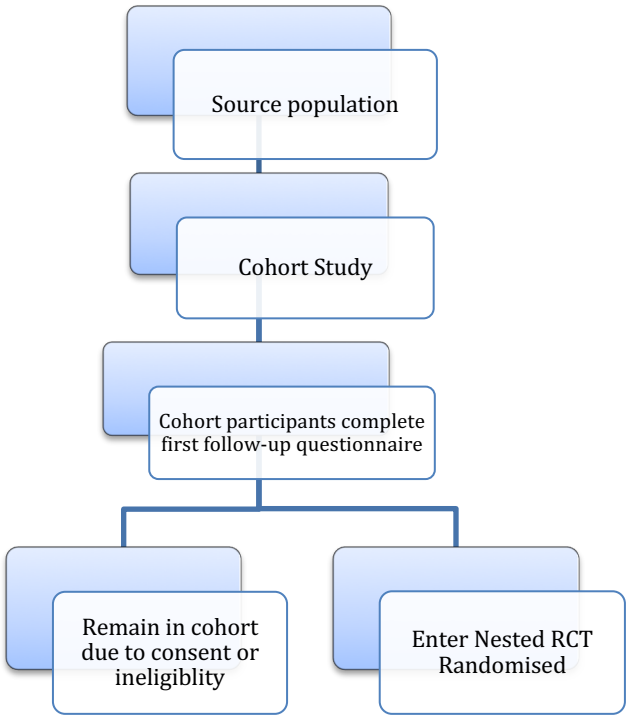

Figure 3.4: Flow diagram of cohort study with nested RCT

Resentful demoralisation can be reduced with this design (Bibby et al. 2018). Enrolling participants into a cohort allows for some participants to be advised and some not to be advised of their randomised allocation. In the case of allocation to usual care (and not to a placebo or alternative intervention), those randomised to this group and who otherwise may experience resentful demoralisation are not advised of their allocation. These individuals know they are continuing in a cohort and may enter a trial at some future time point, they continue unaware they had been randomised to usual care for the trial

and therefore will not be resentful of not receiving an active intervention. Ethics surrounding this approach has been questioned, as participants are not fully informed at each stage of the study, however the methodology provides a resolution to the many issues affecting trials and limits bias within RCTs (Bibby et al. 2018; Kim, Flory & Relton, 2018).

Although rarely used until more recently, the cohort design with nested RCT is starting to be adopted more in the UK and internationally (Kim et al. 2018).

In the UK it was used in a study of transition from child to adult mental health services, allowing a flexible time frame for entering the nested RCT (Singh et al. 2017). It was used in the Yorkshire Health study cohort, which allowed for participant recruitment to 22 RCT's (including studies for first aid, depression and herbal medicine) from the cohort (Holding et al. 2017; Viksveen, Relton and Nichols 2017; Deakin and Tsai, 2017), and it was used within the York trials unit for RCTs within a falls cohort study (Rodgers et al. 2018; Cockayne et al. 2017).

Trials within Cohorts have also been completed in Australia (LBP), Canada (HIV. Mental Health, Scleroderma and Diabetes), Finland (Meniscal injuries) and the Netherlands (Cancer studies), so it is gaining in its popularity (Relton, 2016).

There are few disadvantages to the nested cohort design, but critiques of this design have been conducted. They have predominated around the concern of ethical issues surrounding recruitment and consent to this trial design (Kim, et al. 2018; Biddy et al. 2018; Weijer, Goldstein and Taljaad, 2017; Relton et al. 2016) and suggestions and adjustments have been reported and implemented (Vickers et al. 2018; Weijer et al. 2017; Relton et al. 2016).

Though the design is useful for prevalent conditions it would not be appropriate for incident conditions, as participants afflicted with a condition are selected from a cohort for the nested trial. Investigating incident conditions like fractures would be inappropriate with this design as a cohort would need to be extremely large for sufficient participants to suffer from a fracture and be eligible for any nested trial. The large numbers and extended

time period required to achieve an effective sample size would make the design unworkable for incident conditions.

It is possible that recruitment to a cohort may prove more problematic than a routine RCT and the actual take up of the RCT by the cohort participants may be reduced. The additional cost of following up the cohort as well as the RCT participants may also be a consideration, due to the increased time, effort and cost required comparative to a trial with a run-in period (Dascanio, et al. 2014), however the administrative requirement to collect data from the cohort could be designed to minimise cost and time. Greater costs may also potentially be incurred to encourage response rates of participants to a cohort study comparative to a 'standard' RCT (Boyd et al. 2015).

Research has not currently been conducted to assess the fidelity of this design, to determine if and when this design should be used in preference to a 'standard' RCT (Relton et al. 2015).

3.4.8 Strengths and Weaknesses of RCT designs

A summary of the strengths and weaknesses of the RCT designs discussed are presented within the table 3.2.

RCT Design	Strengths	Weaknesses
General strengths and weaknesses true to all RCT designs	Gold standard methodology Enhanced internal validity Abolishes selection bias Manages temporal changes Manages regression to the mean Provides basis for statistical inference Eliminates Hawthorne effects	Limited external validity Bias can be introduced Perceptive to attrition and resentful demoralisation (without design input) Perceptive to outcome ascertainment (without design input) Time, effort & cost required for design
Placebo RCT	Eradicates placebo effect High internal validity Increased reliability of trial results Eliminates ascertainment bias & Eliminates resentful demoralisation (blinding) Eliminates Hawthorne effect	Limited application to clinical practice Limited for complex interventions Difficult to develop a true and effective placebo for complex interventions High cost
Pragmatic RCT	Evaluate intervention in real world setting Greater external validity and generalisability	Reduced internal validity Negative results cannot infer treatment effectiveness under optimal conditions Challenging design & analysis Greater time and resource required
Randomised crossover trials	Requires fewer participants Participants act as own control Less variable results, paired data comparison Effective for pharmaceutical trials	Carry over of treatment effect Washout period increases trial term, ethical issues - non-treatment period Unsuitable for complex interventions
Pre-randomised consent **Or Zelen's method**	Eliminates resentful demoralisation Reduces attrition	Cross over of participants = attrition Risk of dilution bias, reduced power if follow ITT and crossover occurs Ethical concerns
Patient preference trials	Eliminates resentful demoralisation Offers participant choice to limit cross over	Reduced external validity due to selection bias, with participants choosing not to be randomised Limited comparison across trial arms
Factorial RCT design	Comparison of multiple interventions, combinations to each other & UC Reduced cost	Complex design & analysis required Underpowered for interaction tests

Run in or time periods pre-randomisation trials	Reduces attrition & selection bias Enhanced validity of outcome	Overestimate of population effects Participant loss of interest - wait period
Adaptive cohort design	Design flexibility and adaptability Compares active treatments Introduces and withdraws treatment if superior, inferior or futile. Combines research designs	Potential for over estimation of results due to regression to the mean Costly over extended period of time and with over recruitment Complex design
Cohort design with nested RCT	Reduces attrition & thus sample size Reduces resentful demoralisation Randomise only compliant participants Useful for prevalent conditions	Increased cost with cohort follow up Recruitment to a cohort may be problematic Reduced take up of treatment Inappropriate for incident conditions

Table 3.2: Strengths and weaknesses of RCT designs

3.5 Trial designs and their application to a study of acupuncture and manual therapy for low back pain

Selecting the most appropriate trial design for an RCT is imperative to providing confidence in the results of a trial. As discussed in chapter two, research into acupuncture and manual therapy has been criticised for a lack of quality and unconvincing results, the poor quality of the trials in fact sighted as the reason for not recommending the intervention in clinical practice guidelines (NICE 2016; Liu et al. 2015).

It is imperative that consideration is given to trial design and execution. Acupuncture and manual therapy as complex interventions, require a design to accommodate their complexity, assess them with impartially and evenly, provide adequate analysis, and thus produce a high quality trial, with evidence of effectiveness. In the following sections I shall utilise the information discussed previously for trial designs (section 3.4) and apply them to a potential study of acupuncture and manual therapy.

3.5.1 Placebo RCT and Complex interventions

Placebo controlled trials are excellent designs for testing medication when a placebo pill can be given (Kaptchuk, 2000). Offering a placebo pill in drug trials is relatively straight forward, establishing an inert placebo however has proved problematic for more complex interventions when a therapist interaction is involved (MRC, 2008). Placebo effects are the responses surrounding the intervention and have been shown to have considerable effects on clinical, physiological and brain outcomes for individuals (Wager and Atlas, 2015).

The investigative difficulty in research lies in where the effect size of various placebo interventions differ significantly, resulting in varying estimations of effect of the studied intervention, thus a placebo with a higher effect size would potentially underestimate the effect size of the tested intervention. When comparing the effectiveness of interventions in placebo-controlled studies the effect size and effectiveness of the placebo intervention is a key consideration. Trial results may potentially vary and could be underestimating the true effect of an intervention if studies are comparing the investigated intervention to a placebo / sham with a large effect size. There is variation in the size of effect for different placebos / shams and some may be more active and have significant treatment effects. In recent placebo surgery studies for example, a placebo shoulder surgery was reported to be as effective as the investigated surgery and the results led to it being argued that the investigative surgery should no longer be conducted (Lehtinen et al. 2018). While showing an intervention is no more effective than a placebo may be a very valid undertaking, there needs to be confidence the results are a true representation of the tested interventions. In the studies (Paavola et al. 2018; Beard et al. 2018), the placebo surgery was a very similar procedure to the investigated surgery. Therefore large numbers of participants, an appropriate methodology and effective analytical techniques would be required for the sensitivity to detect a difference in effect size between two very similar interventions, (which could potentially present very similar outcomes and effect sizes) and this was not the case in the studies.

In the Paavola et al. (2018) publication, the study was conducted from 2005 to 2015 and involved placebo surgery which anesthetised the participant, arthroscopically entered the shoulder via four arthroscopic portholes and used saline within the joint to investigate anatomical structures. While considered a placebo surgery for the study, it is a routinely used surgical investigative procedure in clinical practice. It is a very involved procedure, which conducted all but the specific component (arthroscopic sub-acromial decompression, ASD) of the investigative surgery, and both surgeries had a positive and similar effect on the long-term outcome of participants at 24 months (Paavola et al. 2018). It may be that the hypothesised additional component of the sub-acromial decompression (ASD) as a mechanism of action was no more effective above the 'placebo surgery' as the results indicate. Or perhaps the placebo surgery may in itself be an effective procedure (and not a true placebo) and thus the study not sensitive or powered sufficiently to detect a difference. The placebo surgery may have a greater effect size than other potential placebos, while the true intervention had an intervention mechanism over and above the placebo surgery (the ASD), if the placebo surgery had a large effect size and similar outcome to the ASD surgery, as mentioned previously, the RCT would require an extensive sample size with appropriate methodological design and analysis to detect any potential differences between the two interventions. However this study was not planned or powered to consider additive effects of the surgical intervention and the interventions were compared parallel in a three-arm trial. The study also had a limited sample size of 59 participants receiving the ASD and 63 receiving the diagnostic arthroscopy (DA), which did not meet the sample size required to power the study. While designed as a three armed trial, the comparison of ASD to exercise therapy was treated as a secondary outcome and reported that though a small significant difference was seen in favour of the ASD the predetermined minimal clinically important difference was not achieved (Paavola et al. 2018).

Beard et al. (2018) conducted a similar parallel three-armed trial between 2012 and 2015 and produced similar results and recommendations. However they reported an attrition rate of 23% (ASD), 42% (DA) and 12% (no

treatment) of individuals not receiving their allocated intervention by six months. While an intention to treat analysis was reported, a dilution bias may have been introduced with the level of attrition and disparity across the groups. The outcome data reported on 88, 93, 84 participants in the ITT analysis, as opposed to 106 (ASD), 103 (DA), 104 participants who entered the trial, but 72, 65 and 64 in the per-protocol analysis. Thus attrition and a dilution bias may have been introduced into the study due to the systematic differences between the groups, with the greatest overall attrition in the DA arm (103 – 93 – 65) and then the no treatment arm (104 – 84 – 64) an overestimation of the intervention group may have been apparent. The sample size calculated at 85 participants per treatment group for the three-armed trial to power the study was also not achieved.

Beard et al. (2018) reported that while both surgical groups showed a small benefit over no treatment, the differences were not clinically significant. They stated the differences between the surgical groups and no treatment could have been the result of a placebo effect (both surgeries) and / or the postoperative physiotherapy. The headline interpretation from both placebo studies is the addition of ASD over DA is not shown to be effective (Lehtinen et al. 2018), however both surgeries were effective over no treatment and this possibly requires further assessment and research.

There are ethical constraints surrounding placebo surgery, which does limit the types of procedures possible in research (Savulescu, Wartolowska and Carr, 2016). Thus using a well-established investigative operation (DA) with a potentially significant effect size as a placebo, in an RCT that is not adequately powered or an appropriate methodological design to assess additive effects may have limited the sensitivity of the RCT to evaluate and provide a definitive explanation and result for surgery.

The placebo RCT design is more difficult to utilise with complex active interventions (as discussed in section 1.6), like surgery, acupuncture and manual therapy, due to the ethical constraints surrounding the placebo and also it often proves difficult to disentangle the inert, psychological and treatment effects (MRC, 2008; Relton et al. 2010).

The difficulty in disentangling the therapeutic effect can limit the ability to decide on an inactive placebo (MRC, 2008) and it is somewhat difficult therefore to develop a true and effective placebo for an intervention. Especially if the boundaries of the treatment specific effects are blurred and a placebo device is not truly inert, this has been true in attempts to provide placebo or sham interventions for acupuncture and manual therapy trials (Lin et al. 2012; Lundeberg et al. 2011) as discussed previously in section 2.8 and further explored below in sections 3.5.1.1 and 3.5.1.2.

With the growing public knowledge of treatments and understanding of research, it is proving increasingly difficult to blind participants or offer placebo interventions, as many patients have an awareness of the appearance and potential effects of a treatment (Ricker et al. 2015). The applicability of placebo control trials is therefore becoming more of a challenge for complex interventions.

In the next section I shall highlight the placebo type of interventions previously attempted for RCTs of acupuncture and manual therapy and provide a commentary on how effective and accepted these were.

3.5.1.1 Placebo interventions for Acupuncture:

Acupuncture is a multi-facetted therapeutic intervention (Makary et al. 2018). Establishing an effective placebo or sham control for acupuncture trials remains a major methodological challenge. Many attempts to develop effective placebos / sham controls for the investigation of acupuncture have been made, though none thus far have proved to be conclusively inert or superior (Makary et al. 2018; Wang et al. 2017; Zhu et al. 2013).

In their consideration of research for acupuncture, the NICE (2016) LBP guideline committee included many studies using sham techniques including both needle penetrating and non-needle penetrating techniques (section 2.7.2). The results of these studies and the lack of a MCID of 1.0 on a VAS scale of acupuncture over sham acupuncture contributed to their decision not to recommend acupuncture for LBP as discussed in section 2.8). However

many of the included studies had significant methodological issues, were poor quality, with a high risk of bias, but were nonetheless included in NICE's meta-analyses of evidence, apart from Marignan (2014). Table 3.3 summarises the sham acupuncture studies included by NICE (2016), these studies are detailed further in the text below.

Author Year Participants	Intervention	Sham intervention	Outcome Measure / Results
Marignan 2014 12	Auricular (ear) acupuncture One off treatment, immediate follow-up post treatment	Acupuncture needling delivered at non-acupuncture points in the ear	VAS – Pain Acup mean = -0.6 (95% CI, 1 to -3) Sham mean = -4.3 (95% CI, -1 to -6)
VAS et al. 2012 275	Acupuncture Five sessions, over two weeks	Sham = acupuncture needling at non-acupuncture points Placebo = pressure with semi-blunted needle with guide tube at acupuncture points	RMDQ – relative risk to conventional therapy Acupuncture = 5.04 (95% CI, 2.24 – 11.32), Sham = 5.02 (95% CI, 2.26 – 11.16), Placebo = 2.57 (95% CI, 1.21 – 5.46)
GERAC trial Haake et al. 2007 1162	Acupuncture Ten sessions, over 6 – 9 weeks	Needled avoiding acupuncture points or meridians bilaterally	Von Korff Chronic Pain Scale Inter group difference: Acup vs Conventional therapy 20.2 (13.4 to 26.7 P< 0.001) Acupuncture vs sham 3.4 (-3.7 to 10.3. P = 0.39) Sham vs Conventional therapy 16.8 (10.1 to 23.4. P<0.001)
Kwon et al. 2007 50	Acupuncture 12 sessions, over four weeks	Needles insert into non-acupuncture points (10-20mm from acupuncture point)	VAS – Pain Acup baseline = 52.24±19.76, FU = 33mm (SD 15.75) Sham baseline = 51.28 ±21.42 FU = 35.52mm (SD 15.22) not statistically significant
Brinkhaus et al. 2006 301	Acupuncture 12 sessions, over eight weeks	Non-acupuncture points, needled bilaterally using superficial insertion	VAS – Pain Acup mean = 28.7mm (+/- 30.3mm SD) Minimal (sham) mean = 23.6mm (+/- 31.0mm SD)

			Wait list mean = 6.9mm (+/- 22.0mm SD)
Molsberger et al. 2002 186	Acupuncture 12 sessions, 3x per week, over four weeks	Needles inserted but applied superficially at non-acupuncture points	VAS – Pain Acup mean = 23mm (SD 20) Sham mean = 52mm (SD19)
Edelist et al. 1976 30	Acupuncture Three treatments, over six days	Acupuncture needles inserted at non-acupuncture points	Global evaluation – responder criteria Acup 46% improvement Sham 40% improvement
Hasegawa et al. 2014 80	Acupuncture Five sessions, no duration information provided	Non-penetrating needles on same acupuncture points	VAS – Pain Acup mean =1.98cm (SD 2.12) Sham mean =3.38cm (SD 2.26)
Cho et al. 2013 130	Acupuncture 2x per week, over six weeks	Semi-blunt needle on non-acupuncture points	VAS – Pain Acup mean =2.78cm (SD 2.32) Sham mean =4.06cm (SD 2.19)
Cherkin et al. 2009 638	Acupuncture Seven weeks	Simulated acupuncture using toothpick in needle tube, tapped and twisted toothpick on acupuncture point	RMDQ –function Individualised acup mean 4.4 Standardise acup mean 4.5 Sham acup mean 4.4 Usual care mean 2.1 (P<0.001)
Kennedy et al. 2008 48	Acupuncture Three to 12 sessions, over 4- 6 weeks	Park Sham device non-penetrating needle but needle and guide tube contacted skin at acupuncture point	VAS – Pain Acup VAS mean = 28.3 ± 5.2, 95% CI, 17.9 – 38.6) Sham VAS mean = 38.9 ± 5.1, (95% CI, 28.5 – 49.2)
Inoue et al. 2006 31	Acupuncture One off treatment	Guide tube tapped onto the skin at the most painful points and acted as though needle was inserted	VAS – Pain Acup mean = -15±9 (SD) Sham mean = -5±9 (SD) (P= 0.02)

Table 3.3: Acupuncture sham and placebo studies in LBP trials (included within NICE, 2016)

Sham acupuncture techniques have included non-point needling (placing needles in non – acupuncture points or in points not specific to disease according to Traditional Chinese Medicine (TCM)) and minimal / shallow

needing (needles inserted superficially). Trials using comparisons of these sham techniques with acupuncture showed relatively modest differences seen in effect sizes (MacPherson et al. 2017c). However these control interventions all include needling the skin and may have had an active treatment needle insertion component in the sham intervention, as evidence has shown there is a physiological effect on the innervated tissue when it is needled (MacPherson et al. 2017c; Chen et al. 2017; Cummings, 2016; Zhu et al. 2013; Lin et al. 2012; Moffet, 2009; Lund, Näsland and Lundeburg, 2009; Campbell, 2006). In Japanese acupuncture for example, many types of the acupuncture technique employ shallow insertion of needles, of varied location or also non-insertion techniques (toya hari method) (Birch and Felt, 1999). Understanding the physiological mechanism of acupuncture in investigative studies is crucial before an effective study can be designed or an appropriate sham control intervention.

Summary of sham (needle insertion) acupuncture studies for LBP (also see table 3.3*):*

Marignan (2014) trial conducted a one off treatment of auricular (ear) acupuncture for low back pain versus placebo acupuncture (selecting acupuncture points in the ear that did not correspond with low back pain points). 12 participants were split evenly between the two groups and assessed before and after the intervention on the VAS scale. They reported the acupuncture group had an average decrease of pain of -4.3 (-1 to -6) and reported this result as highly significant $P<0.002$ and the placebo group experiencing a decrease of 0.6 (1 to -3. $P>0.28$). The study compared auricular acupuncture points, which corresponded with LBP in TCM principles, with auricular acupuncture points, which did not correspond with acupuncture points in TCM principles. Thus they were comparing acupuncture with acupuncture, just different prescriptions. It is interesting the sham treatment is interpreted as a sham intervention and also to observe the significant difference between the two groups, however the small sample size was very small within this study, it included only a single treatment of acupuncture, and conducted a before and after analysis, the methodology of

this study was very low quality, with a very high risk of bias, thus limiting the interpretation of the results.

Vas et al. (2012) conducted a multi-centred trial with 275 participants across four arms. Acupuncture, sham acupuncture (acupuncture needling at non-acupuncture points), placebo acupuncture (pressure with semi-blunted needle with guide tube at acupuncture points) and conventional treatment. The participants received five treatments across a two-week period, and the primary outcome was reduction in RMDQ score of 35% or more, two weeks after the interventions and the acupuncture interventions were measured as relative risks to conventional therapy. Acupuncture = 5.04 (95% CI, 2.24 – 11.32), Sham = 5.02 (95% CI, 2.26 – 11.16), Placebo = 2.57 (95% CI, 1.21 – 5.46). They reported all three acupuncture interventions were superior to conventional care but there was no significant difference between three acupuncture groups, and reported the results imply acupuncture was not better than sham or placebo acupuncture. The study was well conducted with a low risk of bias. While all the acupuncture interventions showed improvement over conventional care, no difference was found between the interventions and this may be explained by acupuncture merely being a good placebo treatment and any effect seen simply that of a placebo effect (Lowe, 2017; Colquhoun and Novella, 2013). Or alternatively it may be explained in part by the adequacy (or inadequacy) of the interventions used. Five intervention sessions over a two week period may not have been sufficient to establish a treatment effect versus the context of a treatment effect and it also could be viewed as minimal for the treatment of LBP, for example the NICE guidelines (2009) recommended ten sessions over a ten to twelve week period. MacPherson et al. (2013) reported the treatment dose and duration of treatment were shown to effect treatment outcomes, with lower dose and duration showing smaller treatment effects, and this view also supported by a review of 24 chronic pain studies (Chen et al. 2019). In addition the use of the RMDQ as the primary functional outcome measure may have limited the sensitivity of the results presented, as it may be limited in its ability to measure significant functional change over a short period of time, and the RMDQ is reported to be a less sensitive measure if disability is likely to

remain high in a trial (Deyo et al. 1998). In studies of acute conditions, with very short treatment periods, a measure of pain e.g. the VAS may provide a more sensitive outcome measure (Gould et al. 2001).

The GERAC (Haake et al. 2007) large trial with 1162 participants compared verum TCM acupuncture with Sham acupuncture (superficial needling at non-acupuncture points) and conventional therapy (medication, physiotherapy and exercise), for ten 30-minute sessions over five weeks. At six months a questionnaire response rate of 47.6% (acupuncture), 44.2 (sham) and 27.4% (conventional), attrition bias would have been introduced due to the high levels and disparity across groups. An intention to treat analysis was used though a dilution bias may have been introduced and the lower response rate in the conventional therapy could exaggerate the effect of the acupuncture groups, thus potentially overestimating their effect. For acupuncture vs conventional therapy they reported between group differences of 20.2% (95% CI, 10.1% to 23.4% P< 0.001) for the acupuncture versus sham 3.4% (95% CI, -3.7 to 10.3% P=0.39) and for sham vs conventional therapy 16.8% (95% CI, 10.1% to 23.4% P<0.001). Reporting acupuncture and the sham acupuncture were almost twice as effective than conventional therapy. The study was well conducted with a low risk of bias. The results were quickly interpreted and reported as acupuncture being ineffective and no different to sham (Ernst, 2008) with no discussion of the sham intervention used. Lowe, 2017, and Colquhoun and Novella (2013) support this view and argue that acupuncture is purely a good placebo and any treatment effect seen is that of a placebo effect. They feel this principle was demonstrated clearly in several studies with the sham acupuncture and the true acupuncture effects being closely aligned. Others however, have an alternative argument, that the sham was in fact a milder form of acupuncture and questioned the interpretation of the results (Shih, Costi and Teixeira, 2008; Wand and O'Connell, 2008; Bausell, 2009). This study was delivering two types of acupuncture technique, and both acupuncture and sham (non TCM point) acupuncture performed significantly better than conventional therapy in the study. It is possible our understanding of the mechanisms of acupuncture remains limited, and the study design using TCM principles in research

comparing them to non-TCM needling may be misleading. The act of needling may be the active component of acupuncture (Cummings, 2016; Lin et al. 2012; Choi et al. 2012; Campbell, 2006), and thus when comparing the two, sham acupuncture may have diluted the estimate of the effectiveness of true acupuncture. The need to follow TCM principles or perhaps even the need for a TCM practitioner to deliver acupuncture may thus not be necessary, in research or clinical practice (Dascanio, 2015b).

Kwon et al (2007) conducted a study comparing acupuncture with placebo acupuncture, which involved acupuncture needling on non-acupuncture points, with 50 participants across both groups. Treatment was for 20 minutes three times weekly for four weeks. The authors concluded acupuncture (VAS baseline 52.24±19.76, FU 33mm (SD 15.75) was more effective than the placebo (VAS baseline 51.28 ±21.42, FU 35.52mm (SD 15.22) though the results were not significant. There was limited information on allocation concealment or blinding and the methodological quality was poor for this Korean study and as discussed acupuncturing non-acupuncture points as a form of placebo has been refuted as the sham acupuncture may dilute the estimate of the effectiveness of the true acupuncture, (MacPherson et al. 2017c; Chen et al. 2017; Cummings, 2016; Zhu et al. 2013; Lin et al. 2012; Moffet, 2009; Lund et al. 2009; Campbell, 2006).

Brinkhaus et al (2006) conducted a study of 301 participants, comparing acupuncture for LBP with minimal acupuncture, and a waiting list control. Their minimal acupuncture technique involved using acupuncture needles but placing them on non-acupuncture points and not inserting them as deeply as the acupuncture points in the acupuncture group. 12 treatment sessions were conducted over an eight-week period and the VAS was used to measure pain intensity. At eight weeks the difference between acupuncture and the waitlist control was 21.7mm (95% CI, 13.9 – 30.0mm) and the difference between acupuncture and minimal acupuncture was 5.1mm (95% CI, -3.7 to 13.9mm). Both acupuncture and minimal acupuncture performed significantly better than no treatment, but no significant differences were found between acupuncture and minimal acupuncture at eight, 26 or 52 weeks. The study

was of high quality with a low risk of bias giving confidence in the results. While the minimal acupuncture intervention was not referred to as a sham acupuncture intervention in the study or publication (as the authors determined it as an minimal form of acupuncture) critics were quick classify it as a sham technique, and reported the results as demonstrating no difference between acupuncture and sham acupuncture for LBP (Parsons, 2006). Lowe 2017, and Colquhoun and Novella (2013) argue that no difference between acupuncture and sham acupuncture further demonstrates that acupuncture works principally as a placebo treatment, and any effect seen is simply that of a good placebo effect. An alternative argument, as discussed previously, is that the study delivered two types of acupuncture technique, and both acupuncture and minimal acupuncture performed significantly better than no treatment in the study. It may be possible that the physiological mechanisms of acupuncture are poorly understood, and research studies comparing TCM and non-TCM acupuncture could be a flawed from the outset, due to the misinterpretation that TCM points are the active component of acupuncture, thus comparing TCM with non-TCM in research will potentially fail to show the effectiveness of acupuncture. The sham studies potentially demonstrate that the 'effect of the needling' (rather than the TCM point location) may need to be considered with greater importance in research and clinical practice. Thus future sham studies should possibly not follow TCM / non-TCM principles of acupuncture as this may dilute the estimate of the effectiveness of true acupuncture, and they should contemplate other options for sham techniques.

Molsberger et al. (2002) compared acupuncture combined with conservative orthopaedic management, versus sham acupuncture (needles inserted superficially at non-acupuncture points) combined with conservative orthopaedic management, versus conservative orthopaedic management, with 186 participants across the three groups. Both acupuncture (VAS mean 23mm (SD 20) and sham (mean 52mm (SD19) performed better than with conservative orthopaedic management, and they found a significant difference between acupuncture and sham ($P<0.001$) on VAS, though found no difference in the functional mobility or medication use of participants

within the study. They concluded acupuncture could be an important addition to conservative orthopaedic management, however their study methodology was low quality and the results had a high risk of bias limiting the interpretation of the study outcomes. The sham technique delivered a form of acupuncture treatment and thus the study may have been comparing a true acupuncture treatment with a potentially slightly inferior acupuncture treatment, as opposed to comparing to a sham / placebo technique, thus the sham acupuncture may have diluted the estimate of the effectiveness of true acupuncture. As discussed above, this might indicate that the TCM point specificity may not be the key principle in the mechanism of acupuncture, or alternatively it may be that the effect seen in the acupuncture group is merely that of a placebo effect (Lowe, 2017; Colquhoun and Novella, 2013).

Edelist, Gross and Langer (1976) highlighted the use of Sham needling in their study over 40 years ago. The methods and analysis were limited in this study when compared to the standards of RCTs conducted today, but demonstrated the historic principle of trying to address the issue of placebo / sham within studies. 30 participants were enrolled in the study, and allocated to acupuncture or sham acupuncture, (which involved needling 15cm away from the spine at the level of L4-5 and in the posterior leg 10cm below the popliteal fossa). Treatment was conducted three times at two-day intervals. The assessment criteria reported in the study was limited with an orthopaedic surgeon and the patient determining if the patient had improved or not. 15 patients from each group, seven in the acupuncture group improved, 8 did not, six in the sham group improved, 9 did not. They reported a 46% improvement in the acupuncture group and a 40% improvement in the sham group. This was a low quality study with a high risk of bias and thus the results need to be interpreted with caution. The technique used for sham needling demonstrated the belief in the TCM principles of specific point, by just needling at an alternative sight to the true acupuncture group. However, as discussed, many other TCM acupuncture points exist in the region used for the sham technique, trigger point needling techniques would utilise this region with needling to the Latissimus dorsi muscle and as previously stated the insertion of needles creates a physiological healing response. It is possible

the study was actually just comparing two programmes of acupuncture treatment; thus the sham acupuncture may have diluted the estimate of the effectiveness of true acupuncture, and that TCM point specificity is perhaps not the causal treatment principle of acupuncture techniques. An alternative view is that the study may be highlighting acupuncture as simply a good placebo treatment (Lowe, 2017; Colquhoun and Novella, 2013).

Summary: The studies using sham acupuncture with needle insertion provide information suggestive of a limited difference between acupuncture and sham needle acupuncture for low back pain, although probably more noteworthy was that both interventions improved comparative to usual care in most studies.

It maybe that conventional care is inadequate, and thus both interventions present as effective. However as discussed, the consistent finding of an apparent lack of difference between acupuncture and sham acupuncture is interpreted by some as meaning that acupuncture is just a very good placebo intervention and not an effective treatment at all (Cummings et al. 2018; Lowe, 2017; Colquhoun and Novella, 2013). This is view is in part due to the expectation in research to see a dose response of assessed interventions, i.e. true acupuncture would present as better than placebo acupuncture, and placebo acupuncture would present as better than no treatment, if in fact true acupuncture was more effective than the placebo effect. An alternative view is that the sham acupuncture techniques used are not inert or a true placebo, but actually an active form of acupuncture treatment (Cummings et al. 2018; Cummings, 2017; Wang, 2017; Lundeberg et al. 2011; Moffet, 2009; Lund et al. 2009; Lund and Lundeberg, 2006). Kamper et al. (2013) argues that only when it is understood what elements a sham intervention actually control for; can it be established what is accountable for a between group difference in treatment effect. Acupuncture may just simply be a very effective placebo intervention, however further rigorous assessment and research is required to explore this.

MacPherson et al (2017c) argued that the appropriateness of many previously published sham controlled acupuncture studies, which used needle

penetration, were questionable. He suggested the results of the studies he reviewed, indicate sham acupuncture (using penetrating needles) is not inert but an active, and influences the effect size seen in the studies, and thus concluded these sham techniques should not be employed when controlling for non-specific effects of acupuncture (MacPherson et al. (2017c). Whereas Colquhoun et al. (2013) argued that acupuncture is merely a theatrical placebo, with no meaningful differences seen between acupuncture and sham acupuncture. He questions whether the differences that may be seen in effect size may be the result of bias and lack of blinding in the studies, and stated that any benefits which may exist are simply too small to be clinically helpful to patients. Alternatively, while Western Medical Acupuncture evolved from TCM it has adapted and no longer adheres to the Chinese principles, considering that acupuncture needling stimulates the nervous system and stimulates the body's own pain relieving mechanisms (White et al. 2009). Thus offering an explanation as to why needling at non-TCM acupuncture points as a form of sham / placebo may not be demonstrating a significantly different effect in some research to the true acupuncture (Cummings, 2016).

Some acupuncture points have been demonstrated to have greater point specificity on brain imaging studies (Zhang 2004), however great controversy remains over whether the specific location of Traditional Chinese Medicine (TCM) acupuncture points are in fact specific or accurate (Campbell, 2006). While some in the acupuncture community truly believe in true TCM point specificity, it is increasingly being demonstrated that the 'act of needling' may potentially be the active component of acupuncture not necessarily the location of the needle (point specificity). Evidence suggests that the insertion of a needle anywhere in the body will elicit a physiological and potentially a treatment effect (Cummings, 2016; Lin et al. 2012; Choi et al. 2012; Campbell, 2006). Further, Langevin and Yandow (2002) mapped TCM acupuncture points with the anatomical connective tissue planes of a human body and discovered over 80% of the acupuncture points corresponded with the location of inter-muscular or intramuscular connective tissue planes. They concluded the anatomical relationship between the two was relevant in explaining acupunctures mechanism of action. This theory was supported by

a study demonstrating a reduction in the electrical impedance of needled tissue on anatomical connective tissue planes, which corresponded with some acupuncture points (Ahn et al. 2005). Also by a review literature proposing the acupuncture mechanism occurs through the sustained stretch of tissue by the manipulation of the needle, impacting upon the connective tissue plasticity and modulation of the peripheral sensory nervous system (Langevin, 2014).

Western acupuncture, which is the most commonly practiced by physiotherapists and doctors in the UK focuses more on a physiological approach to acupuncture, in its specific effect on the connective tissues (langevin et al. 2007), the body in producing natural pain relieving chemicals, such as serotonin, melatonin and endorphins (da Silva et al. 2014; Wang et al. 2014; Wang, Kain and White. 2008; Stein et al. 2001) and its anti-inflammatory effects by stimulating vagal modulation (Jin et al. 2017; Cummings, 2016; Lim et al. 2016; Torres-Rosas et al. 2014).

The 'act of needling' being the active component is a view potentially also supported by other areas of medical research. Benedetti (2008) reported that the act of inserting an epidermic needle without injecting any fluid, in post-operative patients had a similar pain relieving effect as injecting specific pain relieving medication. Indicating the act of needle insertion may potentially be the active component of treatment. He also repeated this study with the injection of saline, and reported a similar outcome (Benedetti et al. 2009).

Opioid (pain relieving) receptors involved in pain modulation have been shown to exist outside of the central nervous system and on the peripheral sensory nerves. Stimulation of the peripheral nerve opioid receptors will therefore produce an analgesic effect. The process stimulates the release of pro-inflammatory neuropeptides (e.g. endorphins) and the down regulation of neural excitability, resulting in pain reduction. Environmental stimuli will stimulate the opioid receptors to produce an analgesic response (Makary et al. 2018; Campbell, 2006; Stein et al. 2001). The insertion of a needle into the skin therefore will stimulate these peripheral nerves, a-delta and a-beta (responsible for detecting pain, pressure touch and movement) which are

located in the skin, producing a pain modulation response. The local trauma or pressure of a needle induces nerve signals locally, which are also sent to spinal and central levels. Therefore many argue that using shallow or non-TCM point needling would not provide an inactive placebo for research investigation (Chen et al. 2017; Cummings, 2016; Zhu et al. 2013; Moffet, 2009; Lund et al, 2009; Lund and Lundeberg, 2006; Campbell, 2006).

The sham studies using needle penetration all exhibit significant limitations in their design methodologies, substantial heterogeneity of the dose of treatment, style of acupuncture and type of sham needling is apparent across the studies. Many of the studies had small sample sizes, with low quality and high risk of bias, thus limiting the interpretation and significance in drawing conclusions from their results. This is suggestive of the need for either, further high quality, adequately powered studies of sham RCTs, or further investigation and understanding of TCM point specificity of acupuncture and the design of adequate sham techniques in research studies.

Summary of sham (non-needle insertion) acupuncture studies for LBP (also included by NICE 2016):

Toothpicks have been used as an alternative sham to needling and also sham acupuncture needles have been developed in an attempt to find an alternative to placebo needling; devices included the Streitberger, the Park and the Takakure. A device, which uses a stage dagger mechanism, the needle does not penetrate the skin, but a blunt needle contacts the skin and upon a strong tap withdraws into the shaft of a guide tube. These studies are summarised in table 3.3.

Hasegawa et al. (2014) conducted a study of scalp acupuncture (needles in the head) compared to non-penetrating acupuncture needles (blunt retracting needle, blunt needle and shaft of the guide tube in contact with the participant) at the same acupuncture point locations as for the acupuncture intervention for acute LBP. 80 participants were randomised across two groups, and received five sessions and followed up at day 28. The authors concluded scalp acupuncture (mean VAS 1.98cm (SD 2.12) was more effective than sham acupuncture (mean VAS 3.38cm (SD 2.26). However limited information regarding the length of treatment and methods were given in this study, limiting its transparency, quality and potential bias within the study. Worth noting is this was the only study to use scalp acupuncture for acute LBP and thus may be less comparable with studies using body acupuncture. Additionally they consider acute LBP only, which is professed to resolve spontaneously for the majority of individuals (Qaseem et al. 2017) and thus the results may demonstrate an over estimation of the treatment effect.

Cho et al. (2013) conducted a study across three Korean medical hospitals, recruiting 130 participants randomised to real or sham acupuncture (Blunt non-penetrating needles on non-acupuncture points). Treatment was conducted twice weekly for six-weeks. The authors reported acupuncture treatment (VAS mean 2.78cm (SD 2.32) was superior to sham acupuncture treatment (VAS mean 4.06cm (SD 2.19) at eight weeks (P=0.011). However the quality of the study was low with a high potential risk of bias, limiting the interpretation of the results.

Cherkin et al. (2009) conducted a large four-arm study with 638 participants, comparing ten sessions over seven weeks of individualised acupuncture, vs standardised acupuncture (prescribed points) vs Sham acupuncture (Simulated acupuncture using a toothpick in a guide tube, including tapping and twisting the toothpick at the acupuncture point, the tube and toothpick were removed, but then the acupuncture point was stimulated again at ten and twenty minutes with the toothpick placed on the skin and rotated clockwise and anti-clock wise) vs usual care. They reported all groups showed improvement at eight weeks, but the individualised, standardised and sham acupuncture groups all improved (4.4, 4.5, 4.4 points RMDQ) over usual care (2.1 $P<0.001$) and concluded that real acupuncture appeared to be no more effective than sham acupuncture, but both were superior to usual care. Limited information regarding allocation concealment was given for the study and treatment completion rates were 84%, 87% and 90% across the intervention arms, showing levels of attrition, with some variation across the groups. A consideration for this study is the repeated stimulation of the acupuncture point by the sharp tip of the toothpick within the sham acupuncture group. This technique is likely to have created a local trauma to the soft tissues (potentially may have been more uncomfortable than the acupuncture intervention and caused more trauma) and stimulated pain and the peripheral pain receptors/nerves, and thus perhaps may not have been an adequate inert sham technique. Alternatively it may be that the effectiveness acupuncture seen may be simply that of a good placebo effect (Lowe, 2017; Colquhoun and Novella, 2013).

Kennedy et al. (2008) conducted a feasibility study of 48 participants to investigate the feasibility of comparing acupuncture to sham acupuncture (at the same acupuncture points, using the 'Park' sham device, non-penetrating needle, blunt needle contacts the skin and then retracts into the guide tube). Participants received between three and twelve treatments within a 4-6 week period, though no detail regarding any differences in dosage occurring across groups is given. After adjusting for baseline score, pain appeared to reduce in the acupuncture group (VAS mean 28.3 ± 5.2, 95% CI, 17.9 – 38.6) more than in the sham group (VAS mean 38.9 ± 5.1, 95% CI, 28.5 – 49.2) though

statistical significance was not achieved. No significant difference was found between the groups on the RMDQ. As a small feasibility study, it was not powered for the results to reach statistical significance, therefore any inferences from the study need to be made with caution.

Inoue et al (2006) conducted a small study of 31 participants, comparing acupuncture with sham (therapist tapped a guide tube on the patients most painful part and proceeded to act like they had inserted the needle). The participants received one intervention and were assessed before and after the intervention. They reported significant differences with improvement in the acupuncture group with a mean difference of -15±9 compared with the sham group -5±9 (P= 0.02). While positive differences were reported this was a very small-scale before and after study, with limited generalisability, low quality and high risk of bias.

While there appears to be some evidence of a difference between the effectiveness of acupuncture and sham *(non-needle insertion)* acupuncture, the majority of the studies are very small, of low quality, with high risk of bias. The only large scale study (Cherkin et al, 2009) found no difference between acupuncture and sham, though both techniques were superior to usual care and the sham technique using toothpicks was questionable. Further research is required in this area before conclusive understanding of sham and placebo techniques in acupuncture studies can be gained.

A systematic review of literature supports the use of acupressure as a pain relieving technique (a therapist applies manual pressure to an acupuncture point on an individual's skin) as an effective treatment for pain (Pan et al. 2000). Thus a blunt needle pressed upon the skin and a strong tap of a sham needle (with or without some local trauma from the device) and taping a sham needle in place with a plaster, sponge or tape, or applying the pressure of a needle or guide tube on a participants skin arguably provides a form of sustained acupressure (and potentially upon an acupuncture point, which some argue is an integral part of an acupuncture treatment). It has previously been suggested that this contact with a participant's skin has been able to elicit the activation of the peripheral sensory nerves (c-fibres) which are situated in

the superficial epidermal layers of the skin and receptive to pressure, pain, and touch, and thus have a physiological effect on stimulations received (Vrontou, et al. 2013; Lund and Lundeberg, 2006).

While the theory of the development of sham acupuncture needle appears appealing on first assessment, there are reports of the devices often piercing the skin, creating erythema's and the device requiring a strong physiological tap upon an acupuncture point which could be synonymous to acupressure (a healing technique utilises pressure rather than needles). It has been argued that these devices are not inert in their effects (Cummings, 2017; Moffet, 2009; Lund, et al. 2009). A systematic review and meta-analysis considered the appropriateness of both sham and placebo acupuncture by comparing them to other routine treatments. They found both statistically significant differences between sham or placebo acupuncture compared to routine care and waiting list (standardised mean difference for VAS -0.36 (95% CI -0.54 to -0.18, and the chronic pain scale -0.35 (95% CI -0.49 to -0.20). They concluded sham and placebo acupuncture are more efficacious for pain relief than routine or waiting list care and their appropriateness and use for acupuncture research was premature. Recommending the evaluation of sham and placebo interventions as control methods to determine their specific effect and if acupuncture has a specific effect over them. Guidelines to aid future research and standardisation are advocated (Xiang et al. 2018). A systematic review and meta-analysis supported this view concluding these devices had not been rigorously tested, with questions surrounding blinding credibility and the variation in effect sizes seen across trials. They reported the non-penetrating needles were not adequate as inert controls for acupuncture trials and stated further research and development was required (Zhang et al. 2015).

MacPherson et al. (2017c) in their meta-analysis of 29 studies reported that acupuncture was significantly superior to sham regardless of the sham intervention used, they also reported that studies with sham penetrating needles controls, reported smaller effect sizes for the true acupuncture groups than those studies where the sham comparator did not penetrate the skin, suggestive that the effect size of acupuncture was being influenced by the

type of control method used. Of the 29 studies analysed, 18 compared acupuncture with non-acupuncture controls and 20 compared acupuncture with sham acupuncture. They stated:

> *"acupuncture was found to be statistically superior to all types of control intervention for all analysis ($p<0.001$). Effect sizes were larger for the comparison between acupuncture and non-acupuncture controls than for the comparison between acupuncture and sham acupuncture controls (0.37, 0.26 and 0.15 for comparison with sham controls vs 0.55, 0.57 and 0.42 for comparison with non-acupuncture controls for musculoskeletal pain, osteoarthritis and chronic headache respectively)."*
>
> *(MacPherson et al. 2017c)*

They described greater effect sizes with acupuncture when compared to a sham non-penetrating needle intervention (0.43, 95% CI 0.01 to 0.85) than when compared to a sham needle penetrating technique (0.17, 95% CI 0.11 to 0.23) (MacPherson et al. 2017c). Though conventional levels of statistical significance between groups was not achieved.

More recently however MRI studies have suggested that while both sham acupuncture and real acupuncture appear to be eliciting improvement in pain. Sham acupuncture was suggested to be stimulating the placebo circuitry within the brain, stimulating a response, whereas true acupuncture stimulates the primary somatosensory cortex modulating the body's neurophysiological response (Maeda et al. 2017). Thus potentially explaining why both techniques elicit positive responses. Further exploration of the physiological and neurological processes of acupuncture and thus sham acupuncture are required to allow advancement in this area of study.

In addition, sham acupuncture needles are considerably more expensive than acupuncture needles and while only a minor consideration for the purposes of research, at £425 per box of 100 single-use disposable needles (average of 10 needles per session) comparative to £7.31 for 100 regular acupuncture needles. It could be argued that with the ambiguity surrounding the

effectiveness of sham needles not being an inert placebo, it may be more ethical to use funds to finance alternative treatment comparators within acupuncture trials.

Another technique previously used as a placebo or control for acupuncture was laser therapy (active and inactive) (Langevin et al. 2006). Acupuncture is increasingly more commonplace, and growing as a treatment choice with more than 4 million treatments reported annually in 2012 (Hopton et al. 2012). Within studies participants are likely to know that a laser treatment is not a true acupuncture treatment (lund et al. 2009; Ricker et al. 2015).

While uncertainty remains regarding the true effects of acupuncture across trials, a long-term analysis of trial results suggests the benefit of acupuncture compared to no acupuncture controls remains long after treatment has ceased. In their meta-analysis of 20 studies, MacPherson et al. (2017b) considered the long-term effect of acupuncture on chronic pain and from their central estimate were able to report 90% of the benefit of acupuncture appeared to remain at 12 months comparative to non-acupuncture controls. A 50% treatment benefit was sustained at 12 months comparative to sham acupuncture controls in the analysis. Thus reporting a continued benefit of acupuncture beyond the end of treatment time period appears to be evident. They did report significant heterogeneity across their studies and also reported studies including fewer acupuncture treatments across a shorter duration had poorer outcomes, further indicating the need for a more standardised approach across acupuncture research studies.

Without knowledge and explanation of the true specific effects of acupuncture which current research has not successfully provided, it is very difficult to develop an effective inert placebo for the purposes of a placebo RCT. The majority of acupuncture trials lack standardisation of needling location, depth, dosage, retention time, treatment frequency, number of sessions, all of which may have an influence on clinical effect, thus introducing further non-standardised sham techniques in various trials has further fuelled the uncertainty of acupuncture (Zhang et al. 2016). As discussed above and in chapter two (section 2.8), sham acupuncture has, in

some trials, been shown to be as effective as true acupuncture (Lowe, 2017; Zhang et al. 2016; Moffet, 2009). Further research and understanding of the acupuncture mechanism is needed to allow the analysis of the efficacy of this technique against placebo or sham acupuncture devices and establish if acupuncture is merely a very good placebo or if in fact sham / placebo acupuncture is not inert in its application.

3.5.1.2 Placebo Interventions for Manual Therapy:

In the field of manual therapy a variety of placebo / control interventions have been used. These have included massage, sham massage, sham manual therapy and sham manipulation. The assessment in some of the trials of the delivery of sham treatments was based on the clinician's opinion 'that they were not providing a full and effective treatment', they perceived they were delivering a sham, or inferior treatment and thus controlling for time and attention (Cambron et al. 2011). Sham or clinician perceived inferior treatments are dubious and subject to bias. Outcome ascertainment would be a potential bias due to the inability to blind the clinician in a pragmatic trial. A clinician may possibly influence the outcome due to their opinions of the true or sham intervention and they may also display inconsistency in their delivery of any sham or true intervention.

While a significant number of studies included by NICE LBP (2016) guidelines compared manual therapy to self-management, usual care or other interventions, many did utilise various sham manual therapy techniques, Table 3.4 summarises the sham manual therapy studies included by NICE (2016), these studies are detailed further in the text below.

Author Year Participants	Intervention	Sham intervention	Outcome Measure / Results
Ajimsha et al. 2014 80	Soft tissue techniques (massage, myofascial release) 24 sessions	Hands placed gently over treatment areas	McGill Pain Intervention = mean outcome 13.1 (SD6.9) Sham = mean outcome 18.3 (SD7.5)
Bialosky et al. 2014 110	Mobilisation Six sessions, over two weeks	Sham = Motion mimicking intervention but differing biomechanically Enhanced sham = Sham technique with instructional set to enhance expectation	Pain sensitivity SMT mean = 18.56 ± 23 (SD) Sham mean = 23.64 ± 28.93 (SD)
Haas et al. 2014 391	Manipulation 18, 12 or 6 sessions, over 6 weeks	Sham manipulation, consisted of light massage	VAS Manipulation mean = 29.0 (20.8 SD) sham mean = 37.9 (20.4 SD)
Senna et al. 2011 93	Manipulation 12 sessions, over four weeks	Manually applied forces with diminished magnitude	Functional disability scale SMT mean = 29.83 (12.49 SD) Sham mean = 33.46 (12.68 SD)
Santilli et al. 2006 102	Manipulation Treatment 5 days per week, for 5 minutes, up to 20 sessions, in 1 month	Soft muscle pressing, similar to manipulations but not following specific patterns or thrusts	Functional disability SMT mean = 29.83 (12.49 SD) Sham mean = 33.46 (12.68 SD)
Hoiriis et al. 2004 192	Manipulation (Chiropractic) seven sessions, over two weeks	Simulated position but light pressure only applied	VAS Manipulation mean = 1.71 (1.88 SD) Sham mean = 2.21 (2.02 SD)
Triano et al. 1995 200	Manipulation 12 sessions, over two weeks	Low force, high velocity thrusts	VAS Manipulation mean = 1.33 (1.59 SD) Sham mean = 2.17 (2.44 SD)
Waagen 1986 29	Manipulation (Chiropractic)	Adjustment of spine with minimal force	VAS Manipulation = +2.3 sham = +0.6

	2 -3 times weekly for two weeks		

Table 3.4: Manual therapy sham and placebo studies in LBP trials (included within NICE, 2016)

Ajimsha et al. (2014) conducted a study of 80 participants in an Indian hospital, comparing a protocoled myofascial release (type of massage) with a sham technique (clinicians placed their hands over the same treatment areas but applied gentle maintained contact. They provided 24 sessions over an eight-week period. They reported mean differences between the groups for the intervention 13.1 ± 6.9 (SD) and the sham 18.3 ±7.5 (SD) on the McGill pain inventory at 12 weeks. No information on randomisation or allocation concealment was provided, limited methodological information is provided, the data from the six participants who dropped out was excluded from the results and ITT analysis was not used. The quality of this study was low, with a high risk of bias.

Bialosky et al. (2014) conducted a four-arm study of 110 participants comparing spinal manipulative therapy (SMT) with placebo SMT (which involved a thrust to the patient but in a neutral position rather than rotated), enhanced placebo SMT (included placebo SMT and commentary of successful technique to be used) all receiving six sessions over two weeks and a no treatment control group. They assessed changes in pain sensitivity via a supra-threshold heat response and observed a mean decrease of LBP 10.27 ± 18.22 (SD) across all participants regardless of their allocation and significant group dependent changes in pain were not observed, the mean differences between groups for SMT were 18.56 ± 23 (SD) and the sham were 23.64 ± 28.93 (SD) at two weeks post intervention, though limited results and explanations were available. The methodological, analysis and results information reported in this study is limited, no information on randomisation, allocation concealment or ITT analysis is afforded, limiting the quality and increasing the risk of bias of the study.

Haas et al. (2014) considered if there was efficacy to the dosage of spinal manipulation, by comparing 100 participants to each of four groups, one

receiving 18 sessions of manipulation (15 minutes), one receiving 12 sessions and one six sessions and a sham group (light massage). All participants received a hot pack before and five minutes of ultrasound therapy after their allocated intervention. The intervention group receiving 12 sessions had the greatest outcome. The mean difference for pain for the manipulation (12 sessions) was 29.0 (20.8 SD) and for the sham 37.9 (20.4 SD). This was a small study and the primary investigation of this study was dose of treatment, not assessing sham and they reflected in their discussion that the light massage intervention was technically a comparison control rather than a true sham intervention.

Senna and Machaly (2011) compared 93 participants across three arms, 12 treatments of SMT over one month followed by fortnightly treatments for nine months, vs 12 treatments of SMT, vs 12 sham SMT (manual forces applied with diminished magnitude with patient in neutral not rotated position). They reported both the SMT intervention groups improved comparative to the sham intervention at a one-month follow up for functional disability with mean outcome for SMT 29.83 (12.49 SD) and for sham 33.46 (12.68 SD). This was a small study, with the sample size across three groups and the primary investigation not assessing sham.

Hoiriis et al. (2004) conducted a three-arm study of 192 participants with sub-acute LBP (4-6 week duration). Chiropractic manipulation and a medical placebo, vs sham adjustments (light pressure applied to the participant in a neutral position, with no thrusts) and muscle relaxants, vs sham adjustments and a medical placebo. They reported improvements in the intervention group with VAS mean for manipulation 1.71 (1.88 SD) and sham 2.21 (2.02 SD). Only 82.8% (159) of participants completed their allocated intervention, though the attrition was relatively evenly distributed across the groups (13, 17, 13 respectively), ITT analysis was not used in this study and there was a high risk of bias in the study. Additionally they considered sub-acute LBP only, which is professed to resolve spontaneously for the majority of individuals (Qaseem et al. 2017) and thus the results may demonstrate an over estimation of the treatment effect.

Triano et al. (1995) conducted a three-arm study with 209 participants, comparing manipulation vs a manipulation mimic (low force, high velocity thrust) performed daily for six days per week for two weeks and a back education programme. At four weeks the mean difference on the VAS was 1.33 (1.59 SD) for manipulation and 2.17 (2.44 SD), and on the functional disability was 10.6 (11.7 SD) for the manipulation group and 14 (11.7 SD) for the sham manipulation group. They provided a very intensive course of treatment, which is unlikely to be available in clinical practice and the sham technique used a thrust technique in this study, which may have had a physiological effect on the participant and perhaps may not have been an inert sham, potentially affecting the effect size of the study. 81.3% of participants completed the study and missing data was identified in approximately 20% of the final follow-ups. While the study appeared to be of moderate size to start with only 47 manipulation and 39 sham participants data was available in the four-week analysis, highlighting significant attrition and the ITT analysis not used. This study had low quality and had a high risk of bias.

Waagen et al (1996) conducted a small study of 29 participants comparing chiropractic manipulation with sham (adjustment of body using minimal force). Treatment was for 2 -3 sessions per week for two weeks, at two-week review using the VAS they reported manipulation mean +2.3 and sham +0.6. Though limited information regarding the results or the methods of the study was provided. It was of very low quality with a very high risk of bias and thus interpreted with caution.

The majority of sham / placebo manual therapy studies are small with low power, and not using ITT analysis, they have low quality and high risk of bias, limiting the interpretation which can be drawn from them. A systematic review of manipulation versus sham, reports some evidence for manipulation having specific treatment effects over sham, though the review cautions taking strong inferences due to the few available studies, small study sizes, no practitioner blinding, high attrition and ITT analysis often not used (Ruddock et al. 2016).

Choosing massage as a placebo or sham intervention is questionable, as it is a manual treatment used by many clinicians including physiotherapists and while there is a limited evidence base, it is a recommended treatment for LBP by NICE (2016 and 2009) therefore if used as a sham or placebo would not be inert and may bias effect sizes of the interventions within the trials. As discussed above for acupuncture, contact with the skin stimulates the activation of C fibres, therefore contact made for adjustments, performing high velocity movements or massaging would not provide an inert placebo or control (Lund and Lundeberg, 2006).

Alternative placebo options for manual therapy have included the use of inactive ultrasound (sound wave therapy) (Bennell et al. 2010), detuned ultrasound or detuned interferential therapy (electrical nerve stimulation), controlling only for time and attention of the intervention (Koes et al. 1992). The use of a piece machinery as a placebo even if deactivated however would be evident to the participant, they would be aware they were not receiving the active manual therapy treatment intervention and potentially be subject to resentful demoralisation. In a study on patients' perception of care of manual therapy and placebo treatments, it was reported that the placebo intervention of a low level laser pad was not a successful placebo comparison to the active treatment groups (Cambron et al. 2011), thus limiting the effectiveness of using it as a placebo intervention against manual therapy.

Further research is required into appropriate sham techniques or placebo interventions for manual therapy to provide further information in this area (Hawk et al. 2002).

3.5.1.3 Summary of placebo RCTs for acupuncture and manual therapy

The use of a placebo RCT would be very difficult to achieve when analysing multiple complex interventions for LBP. Both manual therapy and acupuncture are complex interventions, which the physiologically processes of, are not yet fully understood, there is a lack of knowledge and explanation of their active treatment components, thus designing an inert placebo of either

intervention would be challenging (MacPherson et al. 2017c; Hawk et al. 2002).

The use of the placebo RCT trial therefore, would be unfeasible without further insight into the specific effects of the complex interventions studied. While the credibility of a placebo RCT results are convincing, an alternative robust method is required to substantiate the results of the intended trial for this study. A pragmatic trial would arguably be a more acceptable and ethical choice for research in this area (Relton et al. 2010).

3.5.2 Randomised Crossover trial for acupuncture and manual therapy

Crossover trials are useful for investigating comparative medications for LBP, as the wash out period allows participants to cross from one medication to another. It is unclear however what the wash out effect of acupuncture and manual therapy would be, but it is unlikely they occur quickly, completely or at the same rate.

An acupuncture trial for allergic rhinitis reported the continued improvement of symptoms at six months following acupuncture treatment (Brinkhaus et al. 2008), and MacPherson et al (2017b) synthesised the data from 29 acupuncture trials using a network meta-analysis and reported the effect from acupuncture trials remaining at 12 months. Potentially making a cross over trial with wash out effects unsuitable. No crossover trials comparing acupuncture and manual therapy have thus far been conducted (Rubinstein et al. 2011).

A complex condition, like LBP may be difficult to evaluate under these conditions. If an active treatment such as manual therapy demonstrated effect in the first treatment period and a patient had gained improvement, assigning them to a comparative treatment like acupuncture in a cross over trial would be unwise. The participant's baseline markers would change for each intervention and an additive treatment effect may occur instead of an alternative treatment effect. Consequently this design would not be useful for

the clinical evaluation of manual therapy or acupuncture in the treatment of low back pain for this study.

3.5.3 Factorial RCT for acupuncture and manual therapy

The Factorial RCT design allows for the comparison of more than one intervention with a reduced cost comparative to two separate trials. Analysis of each intervention individually to usual care, to each other and in combination can occur. The design provides a wealth of information and analysis, with the additional knowledge of additive effects and interactions occurring in well-powered studies.

While the factorial RCT is considered a complicated design with a more complex analysis, it is a design that would suit the aims of a study for this thesis, considering acupuncture and manual therapy for LBP, with the added advantage of combining them and maintaining a usual care arm.

3.5.4 Preference Trials for acupuncture and manual therapy

Preference trials are excellent at offering participants choice and eliminate resentful demoralisation thus reducing attrition. However inflated sample sizes are required, as participants with strong preferences, are not subsequently randomised and are not included within the trial analysis.

The need for a greater sample size and increased cost of providing additional interventions make this design beyond the possibilities of the planned pilot study, however recording participant preference is an objective of any future trial.

Zelen's methodology:

As discussed Zelen's methodology is effective at eliminating resentful demoralisation, however if great numbers do not consent to their randomised intervention a dilution effect can occur and crossover of participants is likely with this design, influencing the trial results.

Though this design is not considered appropriate for the planned investigation of acupuncture and manual therapy for LBP, some characteristics of the approach would be very useful, i.e. the controls being unaware of their allocated intervention to prevent resentful demoralisation and limit attrition.

3.5.5 Run-in period or pre-randomisation time period Trials

The reduced attrition and selection bias afforded these trial results greater validity. Improving the validity and quality of trials is a key consideration for interventions, whose history is shrouded with low quality RCTs.

While participant drop out during the wait period may occur during these trials, it occurs prior to randomisation thus providing a pool of participants less likely to drop out and thus a reduction in attrition. The suggested potential of an over estimation of the population effect should not impact the results as it is anticipated that participants who drop out would have nonetheless dropped out at some point during the trial.

Reducing attrition and selection bias is vital in a trial, while not selected, many of the features of this design will positively influence the design chosen for this study.

3.5.6 Adaptive cohort

While comparing active treatments, the adaptive cohort provides design flexibility and combines several research designs, enabling it to adapt to advances in medicine withdrawing and introducing interventions as indicated, thus providing an insightful design. As discussed earlier withdrawing interventions early may provide an over estimation of effect with regression to the mean. This design and analysis are considered unnecessarily complex and costly for the planned study of acupuncture and manual therapy for LBP, and thus was not selected as it does not require such flexibility or adaptation during the study.

3.5.7 The cohort design with nested RCT

As discussed in section 3.4.7, there are many advantages to the cohort design with nested RCT, including reduced attrition, reduced sample size, reduced resentful demoralisation and the randomisation of conforming participants only. Due to design methodology and the benefits afforded, this design has been selected for the planned investigation of acupuncture and manual therapy for LBP. Further details and methodology specific to acupuncture and manual therapy will be detailed below in section 3.6.

3.6 A Cohort study with nested factorial RCT

The distinct advantages and disadvantages of various trial designs were considered for this thesis study prior to establishing the chosen design format.

A 'cohort study with a nested factorial RCT' has been selected for this study of acupuncture and manual therapy for LBP due to its methodological suitability and many advantages.

This is relatively new and novel methodology, and has been suggested as an effective method for its use with chronic conditions such as LBP (Grant et al. 2006). LBP is a prevalent complex, often chronic condition and the planned interventions are complex interventions, therefore a primary objective in selecting the design was robust methods to maximise recruitment, minimise bias and ensure a study of high quality.

By combining two investigative methods, (a cohort study and a nested factorial RCT) an observational study and a RCT, this design can address various issues, limit the introduction of bias and potentially solve many issues experienced and reported in previous RCTs (Bibby et al. 2018; Kim et al. 2018; Relton et al. 2017; Relton, 2009).

Recruitment: The objective of employing of a cohort study with a nested RCT was to utilise its methodology to enhance participant recruitment and retention thus reduce attrition, while maintaining the rigorous nature of the

randomised controlled trial to assess the interventions (Bibby et al. 2018; Relton, 2009) as shown in figure 3.5. Previous studies of acupuncture and manual therapy have experienced significant recruitment and retention issues (as discussed in section 3.3.4).

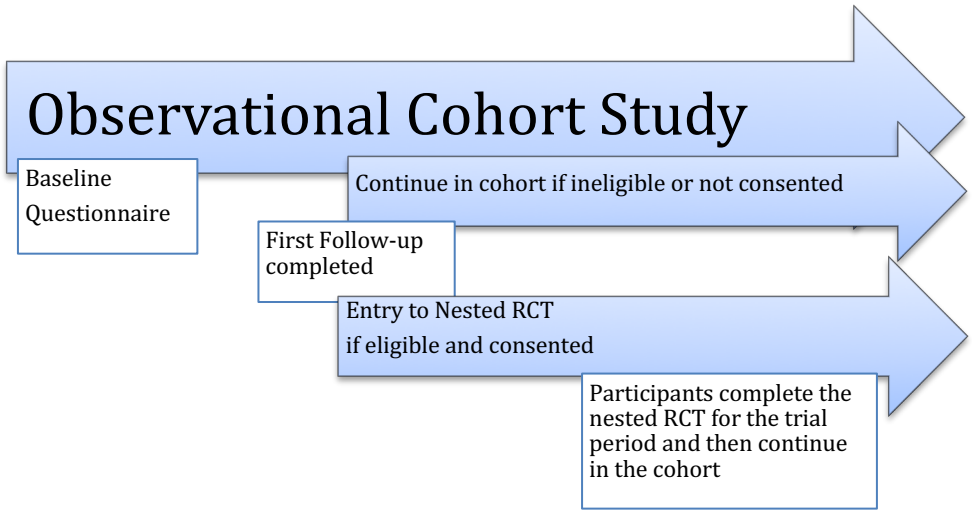

Figure 3.5: Cohort study with nested RCT flow diagram

One of the fundamental and novel benefits of the cohort design is its use as a recruitment vessel to an RCT, maximising potential recruitment. Recruitment initially occurs to the cohort study with advice of potential onward referral to a nested RCT. It is anticipated that having a cohort of participants to recruit for an RCT should prove beneficial, as the appropriate participants are available and have consented to the cohort and potentially the RCT.

The recruitment process incorporates the benefits of a built in 'Run-in period' to the study, ensuring any attrition occurring in the early phase of the study would occur prior to randomisation. It also enables the recruitment and randomisation of conforming participants only, to the RCT. The potential reduction of attrition in the RCT reduces the sample size needed and importantly reduces the risk of attrition bias (Dascanio, 2015d).

Using the design for recruitment and retention also provides participant data across time and a non-RCT comparative group for the study. This approach has not been undertaken previously for a study of LBP.

Additionally this design has enabled the use of some of Zelen's principles to aid the reduction of resentful demoralisation and crossover within the RCT. As participants are within a cohort, those whom are randomised to the usual care arm are not informed of their allocation and thus continue in the cohort, as they were prior to randomisation, unaware they have not been allocated an active intervention arm. Additionally participants will be able to select if they do not wish to receive one of the study interventions but will still be able to be randomised to the preferred intervention or usual care.

Limitations: Some limitations of nested cohort study design do exist, and others will emerge as it becomes more commonly used and will be explored in this thesis.

There are additional costs incurred in following up the full cohort not just those who enter the RCT with a run-in period (Boyd et al. 2015; Torgerson and Torgerson, 2008). Recruitment as for all studies remains a requirement and can be problematic (De Hertogh et al. 2009. Howard and Donovan, 2008), it is unknown if there will be a difference in the uptake to a cohort with a nested RCT comparative to a 'standard' RCT. Additionally participant recruitment from the cohort to the RCT is also unknown, this may prove lower than for a 'standard' RCT and will be explored as part of this study (Dascanio, 2015c).

The ethical implications and consent for this study design remain ambiguous and this full ethical approval will be obtained (Kim et al. 2018; Biddy et al. 2018; Weijer et al. 2017; Relton et al. 2016).

Additionally as this is a pragmatic study, it is not possible to blind the participant or the clinician. The final outcome assessment is provided by the patient using an outcome measure tool but cannot be blinded (Ford and Norrie, 2016; Torgerson and Torgerson, 2008; MacPherson, 2004). Thus blinding is not possible within this study.

Factorial design: The use of a factorial RCT design has been selected for the nested RCT primarily due to the intention to conduct a pragmatic design with complex interventions. The nested factorial design RCT will allow for more than one intervention to be compared, in combination and to a control group.

An extensive evidence report reviewing complementary and alternative medicine (CAM) therapies (including acupuncture and manual therapy) for low back pain, recommended the need for well powered and well-designed trials investigating the use of CAM treatments against widely used active treatments, to help provide clinically relevant outcomes and conclusions (Furlan et al. 2010). Using a factorial design RCT will enable this to occur with acupuncture and manual therapy while also maintaining a usual care control group and allowing their combination.

3.6.1 Addressing bias in a cohort study with a nested factorial RCT

Randomisation eliminates selection bias and the use of an intention to treat analysis both aid the validity of trials, making them more generalizable to clinical practice. Both these principles will be used with the selected design, however there are also other biases that need to be avoided.

Attrition: Attrition is one of the major threats to the internal validity of any trial and acupuncture and manual therapy studies historically have been victim to poor quality studies with high attrition levels (NICE, 2016; Li et al. 2016; Liu et al. 2015). Ensuring low attrition levels to improve the quality of this design is a key objective.

The design of this study specifically aims to reduce that threat (Relton et al. 2010). Using a randomised cohort design will mean a period of three months will occur prior to randomisation, allowing the collection of baseline data and the first set of outcome data. Only participants who return their three monthly data questionnaires (first set of outcome data) will be eligible for the RCT and randomisation.

As discussed, the majority of attrition occurs at the first period of follow-up of an RCT, it is anticipated subsequent attrition following randomisation should be minimal. Dilution effects (i.e. some participants, randomised to intervention fail to accept the treatment) may impact of the study results, however it is anticipated this will be low due to nature of our design and the level of choice provided within the study.

An attempt to reduce attrition will also occur through the provision of comprehensive explanations of the study and regular contact in the form of questionnaires to maintain regular contact with the participants throughout the study (Torgerson and Torgerson 2008).

Resentful demoralisation: The cohort recruitment method will enable the limitation of resentful demoralisation in this study. The design allows for those randomised to the usual care arm not to be advised of their allocation and this is conducted as ethically as possible, and informative explanations will be provided to the participants (Torgerson et al. 1996).

From the outset participants are consenting to enter a LBP cohort study and advised that though they may be approached at a future date to take part in a trial within the cohort, only a proportion of participants may be offered active treatment. Therefore any participants randomised to usual care will not be advised further of any allocation and continue in the cohort as they were prior to randomisation. Not knowing they have entered a trial in the usual care arm aims to reduce any resentful demoralisation.

Outcome ascertainment: Outcome ascertainment bias can be eliminated through blinding, however due to the pragmatic nature of this study blinding is not possible. However using patient reported outcome measures throughout the study eliminates outcome ascertainment bias, as the investigators are not involved in the collection of the participant data (Gupta, 2014; Hróbjartsson et al. 2012; MacPherson et al. 2008).

Hawthorne effects, temporal changes and regression to the mean: Randomisation, the inclusion of a control group receiving usual care and the

inclusion of more than one treatment arm within this factorial RCT will eliminate the effects of the Hawthorne effect, temporal changes and regression to the mean. Any potential effects will be balanced across all four groups (Torgerson and Torgerson 2008).

Participant preference: At initial consent to the cohort study participants will be asked if they are interested in a variety of interventions for LBP. At consent to the RCT participants will be asked if they are willing to receive either acupuncture or manual therapy. If they select, they are happy to receive one intervention but not another, they will still be able to be randomised to their chosen intervention or usual care, allowing them to remain within the randomisation process. Though this does not constitute a full preference trial, it does offer some ability to deselect an unwanted intervention, it aims to limit any participant preference bias, crossover of the trial participants and limit any dilution bias.

3.7 Conclusion

RCT's are considered the gold standard of research and therefore are used to inform the planning and application of clinical practice. There is an inherent responsibility on clinical trials to produce valid and accurate results to inform clinical practice.

There is a distinct lack of evidence investigating acupuncture and manual therapy in comparison or in combination, and though there are studies of the interventions individually the quality of these studies has been questioned and resulted in the case of acupuncture in them being removed from clinical practice guidelines (NICE, 2016). Conducting a well-designed study with high quality and validity was the objective of the planned study for this thesis.

The cohort design with a nested factorial RCT was selected; as it combines the many benefits of previously discussed designs but limits the weaknesses other designs portray. The cohort recruitment design will be used to review its effectiveness in recruiting and retaining participants to a pilot RCT, which

will assess acupuncture against manual therapy, against combined acupuncture and manual therapy, against a usual GP care in a factorial RCT design.

This methodology was applied to enhance participant recruitment and retention while maintaining the rigorous nature of the RCT in assessing the selected interventions. Combining two investigative methodologies and also using the factorial design to allow comparisons and combined interventions will aim to provide a robust informative study.

The cohort study should provide the RCT with eligible compliant participants and reduce the level of attrition post randomisation. The factorial model allows the comparison of alternative treatments in unison and also in combination, while maintaining the usual care group to control for any regression to the mean or Hawthorne effect.

All trials should aim to minimise attrition and trials with superior designs should always be emulated, regardless of complexity. The introduction of bias and errors in research studies can occur through various mediums. Exceptional design and execution of RCTs is required to ensure trials are able to fulfil the expectation lay upon them. Further consideration by the research community of the selection of trial design is essential to limit bias and errors in RCTs.

Gaining further knowledge on the selected design methodology is a key consideration in choosing the design. Finding a way to achieve a reduction in attrition and considering how a run in period may be used without any additional loss through participant indifference will aim to provide greater knowledge to the research community.

Following the consideration and analysis of all RCT designs, the cohort design with nested factorial RCT has been selected for this PhD study. Further information and the full study methods of this novel design are presented in chapter four of this thesis.

4 Detailed Plan of Investigation and Scientific Procedures

4.1 Introduction

In the thesis I have discussed the impact of LBP upon society and I have reviewed and presented the current research and international guidelines, and discussed why there is a need for more evidence for acupuncture and manual therapy for LBP. The quality of previous acupuncture and manual therapy studies were presented to highlight the limitations and weakness of the studies, and I presented the strengths and weaknesses of various trial designs to establish a robust research design to aid future study. A novel trial design was identified and selected, with the objective to realise the need to inform clinical practice and guideline development committees.

In this chapter I will summarise the RCT design and complex interventions previously discussed in chapter three, and outline the justification for a pilot study. I will then provide the detailed methods of the investigation and scientific procedures of the proposed pilot study; a cohort study with a nested factorial RCT of acupuncture and manual therapy for LBP. I will also justify the use of chartered physiotherapists and the selected outcome measurement tools within the study.

4.2 RCT design and complex interventions

As discussed in chapter three, randomised controlled trials (RCTs) have long been acknowledged as the 'Gold standard' of evaluative research method in health sciences (Lilienfeld, 1982; Kang et al. 2008). If wisely designed, conducted and analysed, RCTs provide the strongest evidence for the existence of a cause-effect relationship (Pocock, 1983). RCTs allow understanding of the effects of interventions on participants by means of robust thorough scientific methods (Falagas et al. 2009).

The randomisation process separates the RCT from other forms of research methods and provides high internal validity and allows an RCT to control for selection, subversion or sabotage bias, and through the use of a control arm, it eliminates effect of the Hawthorne effect, temporal changes and regression to the mean. An intention to treat analysis principle also provides the most robust analysis technique for preventing any introduction of bias (Torgerson and Torgerson 2008).

Current treatments for LBP including acupuncture and manual therapy tend to be described as complex interventions (as described in section 1.6 and 3.5.1). The Medical Research Council (MRC) established a framework and guidance for developing RCTs for complex interventions (MRC, 2008). Though the MRC is updating its evidence base on these interventions (MRC, 2018) and this is due to be published in 2019 (Skivington et al. 2018), the adoption of this framework is widely considered to be good practice when researching the development and effectiveness of complex interventions (MRC, 2008).

A working group reviewing complex interventions reported that defining the actual components of a complex intervention is possibly the most challenging part of evaluating a complex intervention (Campbell et al. 2000). Establishing which components are therapeutic and which are inert continues to be open to interpretation, and research currently does not fully translate the replication of or to clinical practice (Moore et al. 2015). The understanding of how complex interventions work within clinical practice is essential to creating a comprehensive evidence base that can direct policy and clinical practice (Moore et al. 2015).

Applying an appropriate research design and defining the components of the intervention where possible is crucial to the methods of the studies (Blackwood, 2006). Standardisation of the process of the RCT and defining which components are fixed and which are variable can assist with assessing a complex intervention with an RCT (Broer, Bal and Pickersgill, 2016; Kühne et al. 2015; Hawe, Shiell and Riley, 2004).

Previous trials of complex interventions for low back pain have routinely compared interventions to usual care or various forms of placebo. LBP is a multi-faceted condition and may require a more complex approach to its management than has previously been investigated. The comparison of treatment interventions to one another and in combination may prove key in developing advancement in the management of this chronic condition. The pilot study methods to be outlined could potentially aid the establishment of preferential treatments and promote the consideration of added benefits in offering additional or combined treatments for LBP (Dascanio, 2015d). In the following section I outline the justification for conducting a pilot RCT.

4.3 Justification for a Pilot RCT

What specifically constitutes a pilot study within social sciences can be difficult to determine, and many full-scale studies conducted within health sciences are relatively small, it has been suggested that many should perhaps be thought of as pilot or feasibility studies (Torgerson and Torgerson, 2008).

The NIHR (2015) indicates that:

> *"Pilot studies are a smaller version of the main study used to test whether the components of the main study can all work together. It is focused on the processes of the main study, for example to ensure that recruitment, randomisation, treatment, and follow-up assessments all run smoothly."*

However in contrast the MRC state:

> *'A pilot study need not be a "scale model" of the planned main-stage evaluation, but should address the main uncertainties that have been identified in the development work.'*

(Craig et al. 2008)

A methodological review of pilot studies within health care from medical journals from 2000 – 2001 was conducted (Lancaster et al. 2004). The

authors concluded no formal description or definition could be established to clarify the specifics of a pilot trial (Lancaster et al. 2004). It is generally agreed however that the principle objective of a pilot study is to facilitate and provide guidance to inform a full-scale RCT (Eldridge et al. 2016). Pilot studies constitute an important role within research, as inadequate preparation and poor methodological process has previously led to poor trial outcomes in studies (Birch, 2004).

A pilot study differs from a feasibility study. A feasibility study often considers if a study can be done. Feasibility studies may not randomise within the trial and are not necessarily intended to evaluate the outcome of interest; they are primarily to provide information on design parameters for a main trial (NIHR, 2017).

Achieving high quality research commences with the selection of an appropriate design of a trial. A design that is able to incorporate multiple facets and situations experienced by those suffering with LBP and that can minimise bias and attrition. Ultimately what is required is a long term, large-scale, multi-regional phase III trial with extensive monitoring of individuals with LBP and their follow-up. It is imperative however to test out the methods of any future full-scale trial, to ensure such methods are effective and appropriate (MRC, 2008). A pilot study allows researchers to gain knowledge on the functioning of the trial and which adaptations if any would be required to improve a future full-scale study.

Justifications and indications for conducting a pilot RCT include:

- *Assessment of the design methods to answer the research question*
- *Testing of the scientific methods*
- *Providing information to plan recruitment and assessing recruitment targets*
- *Trialling the randomisation process*
- *Testing of trial documentation*
- *Determining the acceptance and use of the trial interventions*
- *Surveying for the reporting of incidents*

- *Assessment of the appropriateness of the analysis technique*
- *Testing of the most appropriate outcome measure*
- *Establishing trial parameters to inform future trial power calculations*

(Eldridge et al. 2016; Torgerson and Torgerson, 2008; Salter et al. 2006; Lancaster et al. 2004)

The aims (listed below) of the proposed pilot study were in line with the indications for a pilot study (listed above), i.e. providing information on recruitment, determining acceptance of the interventions, assessing the trial design and methods, testing the trial documentation and outcome measures, and establishing the trial parameters for a future trial. The pilot study was therefore indicated as it aimed to address these to inform a full-scale study:

- *Recruitment rates of participants to a cohort study from GP practices*
- *Recruitment rates of participants from the cohort study to a nested RCT*
- *Consent rates and acceptance of the interventions by participants*
- *Acceptance of the combination of interventions by therapists*
- *Attrition rates of participants pre and post randomisation*
- *The use of this design for evaluating a population with low back pain*
- *Advise the selection of an appropriate outcome measurement tool*
- *Informing a sample size calculation for a full-scale study*

(Dascanio, 2015d)

RCTs are notoriously problematic in their execution, often with limited follow-up periods and other shortcomings of pragmatic trials, therefore ensuring the planning and execution is effective via pilot RCTs is imperative prior to a full-scale costly RCT (McDonald et al. 2006).

A cohort study, used as a recruitment method to a nested factorial RCT, has never been used for the musculoskeletal condition LBP previously. As this will be the first time this study methodology has been adopted for this

condition, the pilot study will investigate the employment of the investigative design to determine parameters for a full-scale multi-centred study.

The outcome of this study would look to inform the methodological design of a national multi-centred trial aiming to establish the nature and course of LBP. Ultimately establishing the most effective interventions and a comprehensive treatment pathway of those suffering with LBP within the UK. Aiming to develop a more informed evidence base to support individuals and inform healthcare policy.

The proposed pilot study will evaluate many of the indicated reasons for conducting a pilot study, the aims of the pilot are detailed below in section 4.4 and the results of the pilot are detailed in chapter five.

This pilot study protocol has been published (Dascanio 2015d, see appendix A3), and is referenced where appropriate throughout this chapter.

4.4 Aims and Objectives of the pilot study

4.4.1 Primary aim

The primary aim was to inform a full-scale study by conducting a pilot study to explore the methodological process of undertaking a cohort design study for the recruitment of participants to a nested factorial RCT, investigating manual therapy and acupuncture alone and in combination versus usual care for LBP (Dascanio, 2015d).

4.4.2 Primary objectives

The primary objectives of this study were to ascertain information to inform a full-scale trial in the following areas:

- *Recruitment rates of participants to the cohort from GP practices*
- *Recruitment rates of participants from the cohort to the nested RCT*
- *Consent rates and acceptance of the interventions by participants*

- *Management of the combination of interventions by practitioners*
- *Attrition rates of participants pre and post randomisation*
- *Inform sample size calculation for a full scale study*
- *Evaluate the use of the design for a population with low back pain*

4.4.3 Secondary objectives

The secondary objectives of this study were to ascertain the most appropriate and clinically useful outcome measurement tool for a full trial.

- *Assess use / completion of Roland Morris disability questionnaire (RMDQ) and Modified Oswestry Disability Index (MODI).*
- *Compare the objective measure tools (RMQD and MODI) to assess which would be most appropriate for a full-scale trial.*
- *Assess use / completion of VAS, SF12, EQ-5D and healthcare utilisation questionnaire*

4.5 Justification for the selected outcome measure tools

Patient Reported Outcome Measures (PROMs): PROMs can be generic or disease specific tool and collect 'patient centred' data at varying time points. The results provide an insight into patient opinion of their own condition, the treatment and how it has impacted upon their day-to-day life (Kyte et al. 2015).

A PROM can be defined as:

> *"…any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else."*

(FDA, 2009)

Outcome measurement has been used in varying forms for more than one hundred years, with reports dating back to Florence Nightingale, who was reported to use a very simplified PROM, with a three-point outcome, simply patients categorised as either 'relieved, unrelieved or dead'. Despite substantial technological advances in modern medicine, some would argue that we haven't actually progressed too far in terms of health care measurement (Appleby et al. 2004).

PROMs were initially developed for their use within research field, to provide objective measures and demonstrate effectiveness of interventions. They are viewed as providing reliable, measurable outcomes and are distinctly rooted within the research process. However following a drive to demonstrate clinical and cost effectiveness in day-to-day healthcare provision and inform commissioning, PROMs have also become routine in clinical practice.

Following the completion of this study, the Chartered Society of Physiotherapy (CSP) advocated the use of the EQ-5D-5L as a generic standardised quality of life questionnaire for data collection within the field of musculoskeletal physiotherapy, and audit data has been collected since 2009 within physiotherapy practice in the NHS (Black, 2013). The tool was selected due to its applicability to a variety of health conditions, its widespread use and based on advice from the Department of Health (DoH). NHS England PROMs team were considering the incorporation of this measure across the NHS (CSP, 2014).

Two condition specific outcome measures have been selected for this study. The motivation for using two objective measure tools for LBP, was to consider which would be the most appropriate measure to use in a full-scale trial. Although many objective measurement tools are in existence, there is still not consensus within the research community or within clinical practice which tool, if any, is the most effective for reliability, validity and sensitivity to clinically relevant changes (Longo et al. 2010). Amongst RCTs of LBP the Roland Morris Disability Questionnaire (RMDQ) (Roland and Morris, 1982) and the Modified Oswestry Disability Index (MODI) (Fairbank and Pysent, 2000) are the most commonly used (Morris et al. 2015) and are

recommended without preference by the Cochrane Back Review Group (CBRG).

The planned sub-analysis of the comparison of the two outcome measures (RMDQ and MODI) is detailed in section 4.21.2, with the corresponding results presented in section 5.4.4 and 5.4.5 and discussed in section 6.6.3. Both measures have thus been selected for this study and the key characteristics are summarised in table 4.1.

Questionnaire	Time point reference	No. of questions	No. of response options	Range of score	Improvement indicated by
Roland-Morris Disability Questionnaire	Today	24	One	0 - 24	Reduction of score
Modified Oswestry Disability Index	Not specified	10	Six	0 – 100	Reduction of score

Table 4.1: Summary characteristics of the RMDQ and the MODI

While two, condition specific outcome measures were selected for their comparison in the RCT, it was important to select one measure as the primary outcome measure for reporting of the trial. Therefore the RMDQ was selected as the primary measure, due to its extensive use in LBP research (Longo et al. 2010).

Roland Morris Disability Questionnaire (RMDQ): The primary condition specific outcome measure chosen was the RMDQ, which was selected due to its extensive use within research studies of low back pain. It was developed in 1983; Roland and Morris were reported to state that:

> *"One of the problems in mounting a trial of treatment of back pain is the lack of suitable outcome measures".*

(Roland and Morris, 1983)

They subsequently went on to develop and present the RMDQ LBP measure (Roland and Morris, 1983). They developed the RMDQ questionnaire to be short, simple, sensitive, and reliable, advocating for it to be used in future trials of LBP. Since its development it was subsequently been followed by the development and use of various alternative outcome measure tools, however the RMDQ has become one of the most commonly used LBP measures in research (Longo et al. 2010).

When a low level of disability is anticipated at the end of a clinical trial, it has been suggested that RMDQ is the most useful tool. It has also been shown to be effective in a primary care setting and is very easy to administer over the telephone, which may aid higher completion rates in trials, at minimal cost (Deyo et al. 1998).

The RMDQ is a list of 24 statements of function and movement covering areas of daily living, physical wellbeing and mental health, with a yes or no tick to answer, the participant reports their response to the questions based on how they are feeling on the day the measure they complete the measure. The RMDQ is scored out of 24 with 0 = no disability and 24 = severe disability (BPS, 2019). A score of four or more on the RMDQ was seen an indication of symptoms differentiable from a dysfunctional to a functional state (Stratford and Riddle, 2016).

The RMDQ has been shown to provide a good measurable, valid, and reliable outcome measure of disability related to LBP (Pawar et al. 2017. Chiarotto et al. 2016; Stevens et al. 2016; Bishop et al. 2010; Brouwer et al. 2004; Fritz and Irrgang, 2001). The internal consistency was judged to be good (Cronbach's α = 0.83 to 0.96) (Smeets et al. 2011; Spanjer et al. 2011; Mousavi et al. 2006; Stratford et al. 1996). The test-retest reliability was good with a range of intra-class correlation coefficients (ICC) = 0.83 to 0.93 depending upon the time period between the test and retest (BPS, 2019; Jordan et al. 2006; Brouwer et al. 2004; Davidson et al. 2002; Stratford et al. 1996).

The RMDQ was reported to have a moderate to large correlation with other PROMS, e.g. ODI and the Quebec scale (QS) (Cohen, 1992) and a moderate to large correlation with pain intensity (r = 0.34 to 0.57) to the VAS, EuroQol and the Oswestry questionnaire (Kovacs et al. 2004). Its responsiveness against global perceived effect scales is rated as good (area under the curve = 0.77), and this was comparable with other scales (ODI and QS) (Davidson et al. 2002). It was also reported to have a moderately high correlation when compared to physical measures of active functional tests (0.55 P = 0.001) (Caporaso et al. 2012).

There are some limitations to the RMDQ. It was reported by Frost et al (2008) that the RMDQ was less responsive to overall change in disability than the ODI questionnaire and Davidson et al. (2002) concluded the RMDQ lacked adequate reliability and scale width for the clinical setting, with poorer test-retest reliability over longer time periods. While the RMDQ is reported to be a useful tool if a low level of disability was anticipated at the end of a clinical trial, by contrast it was therefore less sensitive if disability is likely to remain relatively high throughout a trial (Deyo et al. 1998).

Additionally the measure is only applicable to LBP and the design of the measure (completing the questionnaire based on how ones feels today) could also be viewed as a limitation of the RMDQ, as for individuals with fluctuating symptoms, like LBP, this limits the account of the episodic variation of a participant's symptoms. Completing the questionnaire based on how an individual feels on that day only for a condition such as LBP may afford varying and potentially inconsistent results. Participants may be experiencing a particularly good or a particularly bad day when they complete the questionnaire, potentially not giving an accurate estimate of their standard LBP symptoms over time. Which may affect the estimate of effect being diluted or reduced, and could impact upon achieving the minimum entry threshold into a trial, the adequacy of score reported and the accuracy of scores between the study time-points.

Minimum clinically important difference (MCID): Statistical significance in trial results does not automatically imply clinical importance, very small

treatment effects, too small to be meaningful to participants, may be statistically significant in studies with large sample sizes, and contra-laterally large clinically important treatment effects may fail to reach statistical significance if sample sizes are small (Katz, Paillard and Ekman, 2015). The MCID was established to determine if a treatment effect is clinically important (Man-Son-Hing et al. 2002).

While differences of 4 – 5 on the RMDQ were proposed historically (Stratford et al. 1996) in practice much smaller differences are observed in clinical trials. A meta-analysis of large physiotherapy trials of LBP observed mean between group differences of 0.8 and 1.87 points in the RMDQ at three months (Ellard et al. 2017) and in a LBP study of yoga, a difference of 1.48 points lower was observed at six months (95% CI, 0.33 to 2.62 points) and 1.57 points lower was observed at 12 months (95% CI, 0.42 to 2.71 points) on the RMDQ score (Cox et al. 2010). In a review of treatments for LBP including acupuncture, a difference of 2 - 4 on the Roland-Morris disability questionnaire was reported to be significant (Chou et al. 2007). UKBEAM (2004) observed 1.6 points improvement at three months and 1.0 point at 12 months. While they note it is a small to moderate clinical benefit, it was suggested that this benefit in outcome might lead to large economic effect due to the high cost of LBP in society (UKBEAM, 2004).

This pilot study will plan to observe and consider a planned minimum clinically important difference on the RMDQ for a full-scale study. A difference of 1.5 points on the RMDQ was selected as a prospective figure due to the observed differences seen in previous trials.

Modified Oswestry Disability Index (MODI): The MODI questionnaire was selected as a secondary condition specific measure for this study, to compare the results of the two measures and consider it as a potential alternative measure for a full scale study.

The Oswestry Disability index was first developed as a validated questionnaire in 1980 (Fairbank et al. 1980) and was reviewed and modified

to become the MODI in 2001 (Fritz et al. 2001) it was argued to be the gold standard of functional outcome measures for LBP (Fairbank et al. 2000).

It is a self-reported scale of the functional disability of LBP, including ten subjects (pain, lifting, self-care, walking, sitting, sexual function, standing, social life, sleep, travel). Each subject has a question with six multiple-choice options to select as answers, which follow a scale of difficulty. Each question is scored from 0 to 5, with 0 = no difficulty and 5 = maximum difficulty (each score is multiplied by two for a score of 0 – 100). Overall scores range from 0 – 20 = minimal disability, 21 – 24 = moderate disability, 41 – 60 = severe disability, 61 – 80 = crippling back pain, 81 – 100 = either bed bound or exaggerating their symptoms (BPS, 2019).

The Oswestry Disability Questionnaire (ODQ) has been suggested to be most useful in a trial in which disability is likely to remain relatively high throughout a trial (e.g. chronic low back pain) and is more suited to a more specialist clinical care setting (Deyo et al. 1998). It consists of ten questions with six response options to give a range of severity; it is completed to represent a patient's condition at the time of completion (Smeets et al, 2011).

The MODI has been shown to be a reliable measure, with high test-retest reliability (intra-class correlation coefficient 0.84 (95% CI, 0.73 – 0.91) and it was reported it to be one of the most reliable measures (with SF-36 and QS), having appropriate width scale to reliably detect improvement and deterioration in participants symptoms (Davidson et al. 2002) and having a smaller measurement of error compared to the RMDQ (Chiarotto et al. 2016). This finding was supported across the literature with test-retest reliability reported as high, ICC values range = 0.83 to 0.99 and measurements varied in relation to time intervals, with the greater interval between assessments seeing a lower score (Miekisiak et al. 2013; Grotle et al. 2012; Davidson et al. 2002; Fairbank et al. 2000; Roland et al. 2000).

MODI has good to high internal consistency across studies (Cronbach's $\alpha = 0.71$ to 0.90) (Miekisiak et al. 2013; Grotle et al. 2012; Fairbank et al.

2000; Roland et al. 2000) though no difference between the internal consistency of MODI and the RMDQ were seen (Chiarotto et al. 2016).

MODI's responsiveness to detect change ('area under the receiver operating characteristic curve >0.76) was consistently found across all the four studies (Frost et al. 2008; Davidson et al. 2002; Beurskens et al. 1996; Stratford et al. 1994), and Frost et al. (2008) reported the ODI to be a more responsive measure to change in disability overall (large effect size -0.88 to 1.00 (improved group) & moderate effect size 0.61 to 1.16 (deteriorated group) comparative to the RMDQ (moderate effect sizes -0.70 to -0.74 & 0.69 to 1.25) and the Patient Specific Activity Questionnaire (PSAQ) (large effect size 1.08 to 1.31 & small effect size -0.16 & -0.26).

The MODI was also reported to have a moderate to large correlation with other PROMS, e.g. RMDQ and the Quebec scale (QS) (Cohen, 1992). Its performance was measured against the RMDQ and it was found to be a potential alternative measure to the RMDQ for LBP (Chiarotto et al. 2016; Fairbanks, 2014; Vianin, 2008).

There are some limitations to the MODI. It is time consuming to complete with ten questions each having a multiple-choice response, this also adds an element of complexity and thus some participants may require support in completion of the measure. (BPS, 2019; Smeets et al. 2011). The ten functional subject areas included may not be the most significant to the participants, and the questionnaire has no applicability beyond LBP (BPS, 2019). Additionally as with the RMDQ the participants were asked to complete the questionnaire based on how they feel that day, which as discussed for the RMDQ, limits the potential clinical presentation and account of the episodic variation of a participants LBP, which may result in the estimate of effect being diluted or reduced, and could impact upon achieving the minimum entry threshold into a trial, the adequacy of the score reported and the accuracy of scores between the study time-points.

Minimum clinically important difference (MCID): A minimum clinically important difference on the MODI was reported to be 11 points, in their

assessment of LBP PROMS (Lauridsen et al. 2006) whereas a comparison of the MODI and the QS for physical therapy for LBP, reported a MCID of 6 points on the MODI (Fritz et al. 2001).

In a study of acupuncture compared to usual care for LBP a mean difference of 6.1 (3.5, 8.7) points was observed on the MODI (Nicolian et al. 2019). A study comparing back rehabilitation with spinal surgery, a mean difference of 11.75 was reported as minimal detectable change on the MODI (Johnsen et al. 2013), and a study of back schools for LBP observed a reduction in mean difference in the MODI of 13.49 ± 0.59 ($P<0.001$) (Sahin et al. 2011).

Research has identified the minimum clinically important difference to range of between 6 – 13.49 points on the MODI. It is evident that further research is required to establish a recommended MCID for the MODI.

Summary: Both the RMDQ and the MODI were selected for this study to allow their comparison, to review their use in a study for acupuncture and manual therapy for LBP, and to inform a full-scale study which was the most appropriate measure to utilise.

In addition to the RMDQ and MODI, other PROMS (SF12, EQ5D, VAS and a 'Costs' questionnaire) were also selected for this pilot to assess their use, completion and appropriateness for a future full scale RCT. The SF12 was selected as a quality of life measure, the EQ-5D-3L was selected as a measure of health assessment, the VAS was selected as a measurement of pain, and a healthcare utilisation questionnaire was selected as a measure of healthcare utilisation. These additional PROMS are detailed below:

Short Form 12 (SF12): The SF12 is a quality of life assessment questionnaire, consisting of 12 questions, each having five possible answers ranging from poor to excellent. The answers to the questions are weighted to create two scales, providing a mental health composite score and physical health composite score for a participant. It was developed as a shorter version of the SF36 to aid researchers to restrict the length of surveys while still

maintaining the minimum standards of precision (Utah, 2001; Ware et al.1995).

The SF12 has been widely studied and used as a valid measure of health-related quality of life for a variety of populations. The internal consistency for the individual subscales was good at approximately 0.8 or higher (Turk and Melzack, 2011). The test-retest reliability was report by Huo et al. (2018) to be good with intra-class correlation coefficients (ICC) of 0.61 for the PCS component of the questionnaire and 0.57 for the MCS component of the questionnaire and by Turk et al. (2011) to range from 0.71 and 0.89. It was seen to have linear correlations when compared to the SF36 for both the PCS and the MCS components ($R = 0.86$ to 0.93; $P< 0.0001$) for both sensitivity and reliability (Singh et al. 2006) this result was also supported by the findings Pezzilli et al. (2006).

It is reported that the scoring can be quite time consuming and subject to participant error, unless expensive software is used. The measures scales have few ranges of responses limiting the interpretation and scores are multiplied up to 0-100. It is suggested that due to the physical and mental components of the questionnaire being very closely correlated at high scores they should be used with caution, and the concurrent validity of the mental component may only be adequate and thus not valid for use (Turk et al. 2011).

EQ-5D-3L: The EuroQol group, developed the EQ-5D-3L in 1987 as a standardised measure of health status, with the aim of providing a simple, generic measure of health for clinical and economic evaluation (Reenen and Janssen, 2015). The EQ-5D-3L is recommended for 'Health Technology Assessment' by NICE in the UK, and by the NHS PROMS Group; it is influential in its use to inform economic evaluations and can be used for a wide range of conditions (BPS, 2019).

It is a widely used measure; the first section is a qualitative assessment of health, with questions covering five areas of well-being (mobility, self-care, usual activities, pain / discomfort and anxiety / depression). Each area consists of statements of ability of three levels for each question, it is easy to

complete, with the participants place an x in a box corresponding to their circumstance. The second section is a VAS measure of overall health to measure their health status from 0 – 100 (BPS, 2019; Balestroni and Bertolotti, 2012).

The EQ-5D-3L was found to have reasonable test-retest reliability ICC of 0.70 (Fransen and Edmonds, 1999), values of <0.01 to 0.92 (Varatharajan and Chen, 2012) and Cohen's k test-retest reliability range of 0.29 to 0.61 (Luo et al. 2003) and a higher reliability was associated with an improved health status (BPS, 2019). It was shown to be responsive to change (Payakachatt, Ali and Tilford, 2016), have construct validity, and was found to correlate similarly with the SF-36 for musculoskeletal diseases (Picavet and Hoeymans, 2004).

In a systematic review it was concluded the EQ-5D-3L demonstrated evidence of responsiveness for several conditions and was an appropriate measure for economic evaluation and health technology assessment (Payakachatt et al. 2016).

The EQ-5D-3L is less sensitive to change than the SF12 (Johnson and Coons (2017) and for some conditions it demonstrated a mixed level of responsiveness and it was recommended it should be used alongside additional measures for such conditions (Payakachatt et al. 2016). It is often used for economic evaluations, but it has been reported that being a generic measurement it may not be sensitive enough and may not fully cover disease symptomology e.g. fatigue, concentration, appetite (Priedane et al. 2018). Fransen et al (1999) also reported concerns regarding the difference in patient's self-evaluation and the derived societal utility tariffs as a concern, in addition to a lack of discriminative sensitivity for participants with moderate morbidity.

The wording for some questions has also been suggested to be ambiguous e.g. "walk about…." without clarity of distance or location, thus the mid-range responses can be difficult for participants to select (Janssen et al. 2012).

Additionally it asks participants to report 'how they feel today' and as discussed for the previous PROMS, this limits the account of the episodic variation of a participant's symptoms.

Visual Analogue Scale (VAS): The VAS is a pain rating scale and was one of the first PROMS, used by Hayes and Patterson in 1921. The VAS is a validated, subjective measure of pain and is used most frequently as a uni-dimensional measure of pain. It is easy to use, requiring participants to mark or 'x' placed along a horizontal 10cm line to indicate level of pain along the continuum (line) from no pain to worst pain (0 – 4mm = no pain, 5 – 44mm = mild pain, 45 – 74mm = moderate pain, 75 – 100mm = severe pain, however these values are imposed by the assessor and do not necessarily reflect the participants meanings (PRS, 2019; Delgado et al. 2018). The participant completes the questionnaire based on how they feel in the previous 24 hours and is reported to be and sensitive to small changes (Streiner and Norman, 1989).

The VAS is versatile and can be used for a variety of pain conditions. The same scale can also be adapted to measure other variables (mood, sleep, function) and is quick to complete, requiring no training (PRS, 2019).

The test-retest reliability was shown to be good for the VAS with intra-class correlation coefficients (ICC) of 0.97 (95% CI, 0.96 to 0.98) and 90% of pain ratings being reproducible (Bijur et al. 2001), Alghadir et al. (2018) supported this finding also reporting ICC of 0.97 and concluding the VAS had excellent test-retest reliability.

The construct validity of the VAS was good; with it being highly correlated other scales and was also reported to be more sensitive to small changes than other measures (5-point Verbal descriptive scale, Numeric Rating Scale) (Gould et al. 2001; Joos et al. 1991)

The measure is only useful for uni-dimensional measures and is limited to a single subject i.e. pain in the information provided (PRS, 2019). A higher distributional use of the mid-point on the VAS scale has also been raised as a

concern with it being selected 2.8 - 4.8 times more than other points. While it was shown to be less of an issue with a mid-point indicator, radio button or number removed from the scale, there did appear to be a draw of participants to the mid-point (Couper, Tourangeau and Conrad, 2006). Additionally while the VAS appears a useful measure of pain, it was reported not to perform as effectively when used for disability. Its validity for assessing disability was classified as poor due to a weak correlation with other PROMS (SF-36, RMDQ & VAS for pain) (Boonstra et al. 2008).

Health Utilisation Questionnaire: The healthcare utilisation questionnaire is a bespoke set of questions designed to collect information on health care utilization, developed by the York Trials Unit. It aims to give an indication of healthcare utilisation of individuals within the study. It may be a useful addition to a full-scale study to indicate any differences between groups or with the cohort participants in healthcare utilisation throughout the cohort study period.

The included questionnaires will be reviewed for their completion rates and ease of use, to determine their appropriateness and clinical usefulness for inclusion in a full-scale study.

4.6 A Pilot Cohort Study with Nested Factorial RCT design

The pilot study design proposed, follows a novel cohort recruitment design, in which the pragmatic factorial RCT will be nested within a cohort study.

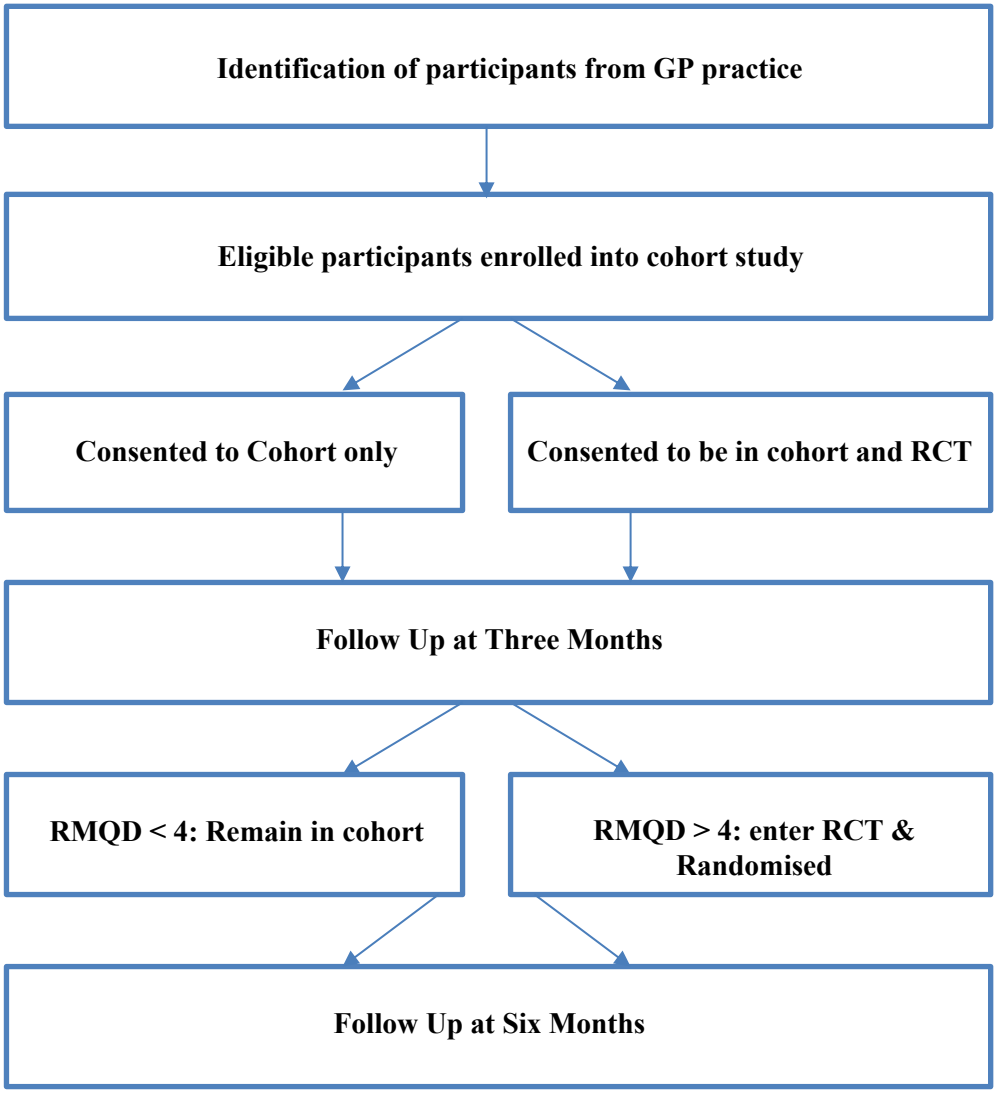

Figure 4.1: Flow diagram of the cohort and RCT recruitment process (adapted from Dascanio et al. 2011)

Participants were recruited to a cohort study and were advised they may be selected for recruitment to a nested RCT during the cohort study period, as detailed in figure 4.1. This allowed for cohort participants who conformed to the study requirements only (returned their three month questionnaires) to be recruited and randomised into the RCT, thus aiming to reduce post randomisation attrition (Relton et al. 2010).

4.6.1 Cohort design

Participants were selected from two GP practices, using coded databases searches. Those individuals aged between 18-65 years who had consulted their GP in the preceding 18 months with low back pain were identified. The study inclusion criteria used during the searches are detailed in table 4.2 and the study exclusion criteria used during the searches are detailed in table 4.3.

All potentially eligible participants were sent an information pack from their GP practice, which contained an invitation letter; a participant information sheet and a study consent form (appendices C4, C5).

It was explained to participants in the information sheet that during the course of the cohort, a treatment trial would occur, and cohort participants would be given the opportunity to take part in the RCT. Those who indicated they did not wish to take part in a future RCT on their consent form would continue in the cohort study only.

All participants who consented to enter the cohort only were went a baseline questionnaire. Those participants who consented to enter the cohort and be contacted about an RCT study were then sent a second participant information sheet, a second consent form and a baseline questionnaire to complete and return to the University of York. Those who returned their completed baseline questionnaire (and consent forms for those consenting to the RCT) were all registered for the cohort.

Throughout the cohort study the individuals were monitored using questionnaires at three monthly intervals. Participants received information explaining that they were forming part of a cohort study and would be monitored regularly (i.e. three monthly) via postal questionnaire, to investigate their LBP over an 18-month period, however this thesis considered participant data to the six-month time point only.

Inclusion Criteria:	Rationale for Inclusion:
Individuals aged between 18 – 65 years	Inclusive of a wide population sample
Individuals registered with a GP practice participating within the trial	Restricted to GPs inducted into the study
Individuals able to converse in English	Limited funds of a Pilot for translators
Individuals who have consulted their GP with mechanical or simple LBP in the preceding eighteen months	LBP reported within the preceding 18 months if still ongoing would be constitute chronic or recurring LBP
Individuals suffering with LBP for between six weeks and eighteen months	Restricted to more than six weeks, due to prior would be considered acute and likely resolve. Limited to an 18 month history, due to increased chronicity changing the presentation of LBP
Individuals with referred pain into the leg will be included in the study (if there was no indication of any serious neurological conditions when assessed by their GP)	Leg pain is common with LBP sufferers and often benign and caused by referral from mechanical structures of the LBP e.g. nerves, muscles, discs.
Individuals with pain present on assessment and pain that is persistent in nature (i.e. occurring at least once daily for eighty percent of the days in the history of their recent painful episode)	Due to the need to differentiate between those with minor and more severe LBP. The condition needs to be present for a substantial period of time to allow the interventions to take place
Individuals who agree to avoid physical treatments other than the study interventions for the ten to twelve week period of the pilot study (active treatment participants only)	Other treatments at the same time as the intervention treatment could impact and alter the participants response
Individuals with a score of four or more on the Roland-Morris disability questionnaire at baseline (UKBEAM 2004)	To enable there to be scope for a change in the outcome measure (RMDQ)

Table 4.2: Study inclusion criteria and rationale (adapted from Dascanio, 2015)

Exclusion Criteria:	Rationale for Exclusion:
Individuals who were below 18 years of age	Children were not considered for this study as they required specialized services
Individuals who were above the age of 65 years	Due to the increased risk of co-morbidities
Individuals with clinical indications of serious spinal or neurological pathology, as assessed by their GP	As would constitute LBP caused by a specific pathology. This study is focused on non-specific LBP
Individuals with a history of spinal surgery	As may alter the clinical presentation of LBP and clinical outcome
Pregnant women or those who have given birth in the last twelve weeks	As this may alter the clinical presentation of LBP clinical outcome
Individuals who had received manual therapy or acupuncture in the preceding three months	As this may alter clinical outcome and may provide an inaccurate measure of treatment effect
Individuals with blood disorders, receiving anti-coagulants or anti-platelets	As this is a relative contraindication to acupuncture
Individuals who are immuno-compromised	As this is a relative contraindication to acupuncture
Individuals with metal allergy	As this is a relative contraindication to acupuncture
Individuals who are unable to provide consent	Due to limitations of this pilot individuals only who can consent to participate are able to be included
Individuals with a history of psychosis or alcohol abuse	Due to difficulty in assessing outcome
Individuals who have a needle phobia, unless individual chooses to participate in the non-acupuncture part of the study	Due to acupuncture being part of the study
Individuals with valvular heart disease or pacemakers	As an absolute contraindication to acupuncture

Table 4.3: Study exclusion criteria and rationale (adapted from Dascanio, 2015)

The cohort study provided a recruitment base for the nested trial, however the cohort participants (not willing or eligible to be part of the nested trial) continued to be monitored in the study and provide cohort data for LBP over the six month period. The data collected from this group of individuals would provide information on a cohort of individuals with LBP who were not included in the trial.

Box 4.1 details a summary of the participant information process.

Box 4.1: Participant information summary

- *A letter invited the patient to join the cohort study (if they are currently still experiencing their LBP) by returning the completed forms to the University of York*
- *The participant invitation letter explained to the participant that there would be a future treatment trial within the cohort and if they wished to also take part in the treatment trial to express that wish on the consent form in the allocated box*
- *Those participants who consented to the cohort and the treatment trial were then sent a second information pack containing the participant information sheet two, consent form two and the baseline questionnaire*
- *Participants who consented to the cohort only received the baseline questionnaire booklet but were not be entered into the treatment trial or randomised*

4.6.2 Nested factorial trial design

Participants from the cohort study were assessed for their eligibility to enter the nested trial; they were required to meet the criteria set prior to commencing the study:

- *Participants who consented to both the cohort and nested trial*
- *Participants who returned their completed consent form and baseline questionnaire*
- *Participants who subsequently returned their three monthly questionnaire*
- *Participants with a score of four or more on the Roland Morris Disability Questionnaire(RMDQ) as the minimum entry requirement*

Figure 4.2 shows the flow diagram of the participants' pathway through the pilot cohort and nested RCT.

Individuals who scored four or more, on the three-month RMDQ questionnaire, were considered to have significant symptoms to differentiate between those with functional or dysfunctional state (Stratford and Riddle, 2016). Symptoms associated with a score of four or more may potentially trigger a referral to physiotherapy via routine contact with their GP as part of a usual care pathway, thus a score four or more on the RMDQ was used as the entry threshold for the RCT.

As shown in figure 4.2, those participants who scored less than four on the RMDQ at three months were excluded from the nested RCT, as their symptoms of LBP were considered minimal and below the threshold to differentiate between those with a functional and dysfunctional state (Stratford et al. 2016). In standard routine care these participants would be managed by their GP with medications and advice, onward referral to physiotherapy would rarely be indicated. These participants were not considered severe enough for further treatment and may well be less responsive to a course of treatment or may be too healthy to demonstrate a true effect of treatment interventions within an RCT. While this study was a

pilot and was not planned to determine a treatment effect, the objective was to keep the trial as similar to an anticipated full-scale trial as possible.

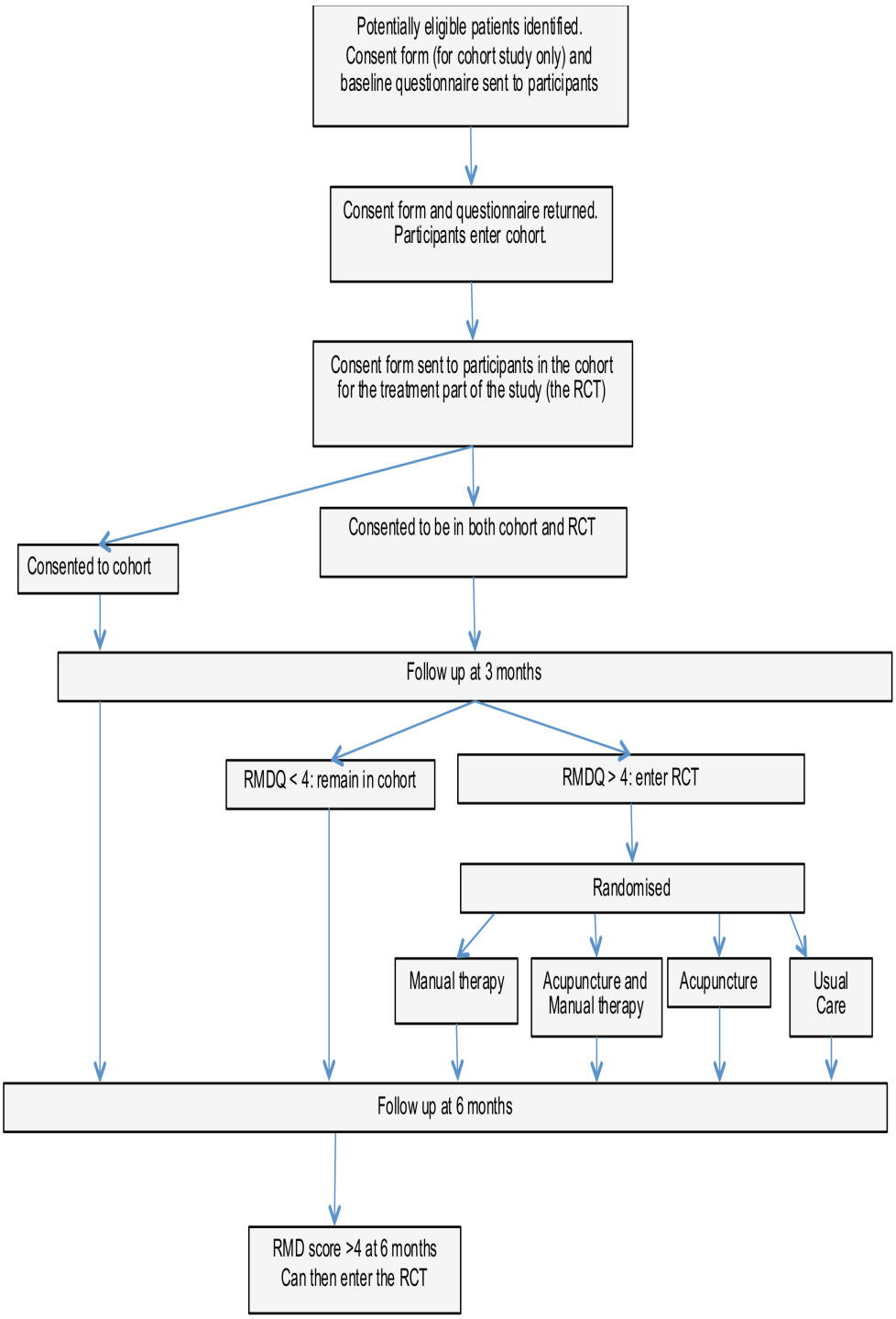

Figure 4.2: Flow diagram of Cohort study with nested RCT (Dascanio, 2015d)

The Roland Morris Disability Questionnaire (RMDQ) was selected as the main treatment outcome measure for this study primarily due to its extensive use within research studies of LBP; as discussed (section 4.5) it has been shown to provide a measurable, valid, reliable outcome measure (BPS, 2019: Pawar et al. 2017. Chiarotto et al. 2016; Stevens et al. 2015; Smeets et al. 2011; Spanjer et al. 2011; Bishop et al. 2010; Brouwer et al. 2004).

The Modified Oswestry Disability Index (MODI) Questionnaire was selected as a secondary measure of back pain, as discussed (section 4.5) to investigate its performance as a measure against the RMDQ and consider it as a potential alternative measure (BPS, 2019; Chiarotto et al. 2016; Fairbank, 2014; Miekisiak et al. 2013; Grotle et al. 2012; Vianin, 2008; Frost et al. 2008; Davidson et al. 2002).

Acceptability and use of the RMDQ and MODI measures were considered by analyzing response rates and results, which aimed to help determine which back pain measure should be the primary measure for a full-scale trial. The specifics of the objective measure tools are detailed in this chapter in section 4.5, discussion of their use and comparative results from the pilot study are presented in chapter five.

A nested pragmatic factorial pilot trial design was used. The factorial trial had four groups:

Treatment 'O' = Usual Care

Treatment 'A' = Acupuncture

Treatment 'B' = Manual Therapy

Treatment 'AB' = Acupuncture and Manual therapy

In this pilot and a potential full-scale trial, this design would allow us to analyse the effectiveness of the two interventions at the same time, while maintaining a comparative control arm, and would increase the efficiency of the trial (Finlay et al. 2003; Byth and Gebski, 2004). In addition, it would allow the investigation of the effects of the two treatments when given in

isolation and compared head to head and if their effectiveness is changed when provided in combination (table 4.4).

This trial design allows more than one research question to be addressed as efficiently as possible without the generation of additional cost of another study (Byth et al. 2004).

The design assumes no interaction between acupuncture and manual therapy and the pilot is not powered to detect such an interaction. If the interaction effects of the two treatments were to be assessed in a definitive trial this would require a significant increase in the total sample size to account for interaction (i.e. 4 times that needed to look at the main treatment effects).

Acupuncture	**Manual Therapy**		
	B (Yes)	O (No)	Margin
A (Yes)	Acupuncture and manual therapy (cell AB)	Acupuncture (cell AO)	All A (cells AB & AO)
O (No)	Manual therapy (cell BO)	Usual GP Care (cell OO)	All non-A (cells BO & OO)
Margin	All B (cells AB & BO)	All non-B (cells AO & OO)	

Table 4.4: Factorial design of pilot RCT

Treatment effects and potential indications of interactions will be monitored, and appropriate clinical outcomes determined to inform an appropriately powered trial, aiming to investigate the effectiveness of the treatments individually and in combination. A full-scale trial would aim to look at the combined treatment effect of acupuncture and manual therapy using the factorial approach.

A pragmatic trial was proposed to allow the interventions to be delivered in line with current clinical practice. It is a more acceptable approach to clinicians and affords greater clinical autonomy, external validity and generalizability with more real life treatments (Relton et al. 2010; Ross et al. 1999).

4.7 Trial interventions

The detailed procedural processes of each trial intervention are detailed below. A TIDieR checklist and guidance (Hoffmann et al. 2014) was used to provide comprehensive and repeatable information.

Usual GP care intervention group (Treatment 0):

Usual care for this study constituted 'usual GP care' i.e. as participants would receive usually under the supervision of their GP or other health professionals as appropriate, and as would be routine in clinical practice, i.e. the same as if they were not involved in the study. All participants received usual GP care.

Usual care also involved the provision of the 'Back Book' (Burton, Main and Cantrall (2002); this is a self-help book for LBP frequently distributed by health care professionals. These participants were not provided with manual therapy or acupuncture throughout the study period. Data was collected on all patients during the cohort period.

Trial interventions:

Information pertaining to all trial intervention groups is detailed below:

All intervention treatments took place at a local private physiotherapy clinic and was provided in face-to-face individual treatments by the recruited Chartered physiotherapists only (see section 4.10), each physiotherapist was appropriately qualified in the techniques they delivered, either acupuncture or manual therapy or both.

Participants were screened and assessed by the chartered physiotherapist using a health-screening questionnaire (appendix C10) and a physiotherapy LBP assessment form (appendix C11) prior to commencing the trial intervention. This was conducted as a second screening process to ensure the participants were safe to receive their allocated interventions and that no red flags of health concerns were present. This process is standard procedure and in line with the professional standards expected of all Chartered Physiotherapists when receiving a new patient to ensure the health of the patient is appropriate for receiving physiotherapy. If any concerns were raised regarding a participant the physiotherapist would contact the trial coordinator to report the concern, discussion between the two would occur and a decision made regarding the participants safety and eligibility to continue in the trial.

All usual standards of care, protocols and practices continued to be observed for all intervention groups. Participants were also provided with usual GP care as required, including the provision of the 'Back Book' (Burton et al. 2002) as would be expected were they not involved in a trial.

Acupuncture intervention group (Treatment A):

Acupuncture has its history in Chinese medicine and is one of the oldest forms of therapies available. It involves the insertion of fine needles into the body and aims to take a holistic (whole body) approach to treatment. In Chinese philosophy illness is considered an imbalance of energy sources in the body and acupuncture strives to recreate this balance and harmony within the body (Vera et al. 2018; Chong et al. 2015). In western medicine acupuncture is considered to stimulate blood flow and nerve activity (Kim et al. 2016; Li et al. 2013; Uchida and Hotta, 2008; Inoue et al. 2005) and stimulate specific areas of the brain to release chemicals that have pain-relieving responses (Jin et al. 2016; He et al. 2015; Bradnam 2011; Lewith, et al. 2005). Specially trained professionals can deliver acupuncture, these include: acupuncturists, physiotherapists, osteopaths, doctors, nurses, and chiropractors. Chartered

physiotherapists only were recruited to this study to deliver the interventions (see section 4.10).

A group of experienced musculoskeletal physiotherapists with acupuncture training were recruited and inducted into the trial. They met prior to commencement, to discuss guidelines for the expected best practice standards and agreed a consensus for the treatment of LBP using acupuncture principles. These included discussion of any other physiotherapy interventions (though this could not include any manual therapy techniques) they would routinely have provided in addition to acupuncture for individuals with LBP, so this may be standardized to all participants. The STRICA guidelines of reporting trials of acupuncture were adhered to (MacPherson et al. 2002).

Participants allocated to this intervention followed a programme of 10 x 30 minute acupuncture treatment sessions, which occurred weekly if possible, over a ten to 12 week period. This was in line with the recommendations from the NICE LBP guidelines (2009).

Acupuncture point selection and needling treatment was delivered as deemed appropriate by the physiotherapist, following the pre-agreed trial guidance (see below) and following their professional governance as required by their professional organisation and standards.

Acupuncture point location, size of needle and depth of needling was applied as detailed in the 'Color Atlas of Acupuncture' (Hecker et al. 2001). It was agreed and anticipated eight to twelve acupuncture needles would be used per acupuncture session for each participant, with a mixture of local and peripheral points used. Points were selected based on their prescription use for LBP and pain as indicated by TCM theory (as detailed in Hecker et al. 2001).

All acupuncture points used for each participant for each session would be recorded in a table by the physiotherapist and documented as part of the participants medical notes at the physiotherapy practice.

Acupuncture point	Bilateral / Unilateral Depth of needles	De Qi stimulated	Time period of needling	Adverse events

Table 4.5: Recording of acupuncture treatments

Achieving the sensation of De Qi (a deep aching) was an objective at each acupuncture point, in line with the theory of acupuncture core teachings, as is described in TCM (Hecker et al. 2001). If a participant experienced any significant discomfort deemed outside normal expectations of acupuncture (i.e. sharp pain at the site of a needle), the needle would be removed, and the participant's sensation documented. This was in line with the routine process of acupuncture treatment in clinical practice by Chartered Physiotherapists. An alternative needle location would then be selected to continue the participant's treatment.

Single disposable needles were used for all treatments; these were package as one needle to one guide tube in strips of five needles. Needles within their expiry date only were used and all needles were disposed of in a sharps bin, in accordance with the physiotherapy practices sharps policy.

It was not permitted for the physiotherapists to provide any form of manual therapy for this intervention group.

Manual Therapy intervention group (Treatment B):

Manual therapy is a form of therapy that involves a therapist's hands to deliver mobilization, massage or manipulation of joints or soft tissues in the body. Specially trained professionals can deliver manual therapy including; physiotherapists, osteopaths, doctors or chiropractors (NICE, 2009) chartered physiotherapists only were recruited to this study to deliver the interventions (see section 4.10).

Participants allocated to this intervention followed a programme of 10 x 30 minute acupuncture treatment sessions, which occurred weekly if possible, over a ten to 12 week period. The NICE LBP guidelines (2009) recommended 9 sessions of manual therapy over a ten to 12 week period but in the interest of equipoise between interventions in the study, 10 sessions were offered for both acupuncture and manual therapy.

The physiotherapists provided the manual therapy intervention as they deemed appropriate to their participant, following the pre-agreed guidance of best practice established for the trial and following their professional governance as required by their professional organisation and standards. The physiotherapists used their hands to deliver the manual therapy interventions to the participants (manual therapy techniques included: massage, joint mobilisation, soft tissue mobilisation, frictions, manipulation, stretching, mobilisations with movement) and all treatments were documented as part of the participant's medical notes at the physiotherapy practice.

If a participant experienced any significant discomfort deemed outside normal expectations of manual therapy (i.e. sharp pain) at the site of treatment, the treatment would be ceased, and the participant's symptoms documented. This was in line with the routine processes of manual therapy treatment in clinical practice by Chartered Physiotherapists. An alternative manual therapy technique would be selected to continue the participant's treatment.

It was not permitted for the physiotherapists to provide acupuncture to this intervention group.

Combined Acupuncture and Manual Therapy intervention group:

A discussion occurred with the recruited physiotherapists to agree the format for providing manual therapy and acupuncture within the same treatment session. The same physiotherapist delivered both interventions in the combined group; they had training in both acupuncture and manual therapy. For the combined manual therapy and acupuncture intervention group, it was

anticipated that the participants would receive a 50% longer treatment session to allow for both interventions to be completed effectively.

Participants allocated to this intervention followed a programme of 10 x 45 minute acupuncture treatment sessions, which occurred weekly if possible, over a ten to 12 week period.

The acupuncture was delivered in exactly the same way as detailed above for the acupuncture intervention group, and the manual therapy intervention was delivered in exactly the same way as detailed above for the manual therapy intervention group, but they were delivered one after the other within the same treatment session.

Treatment was delivered as the physiotherapist deemed appropriate following the pre-agreed trial guidance provided prior to the trial and following their professional governance as required by their professional organisation and standards.

4.8 Study Documentation

All pilot study documentation was subject to the scrutiny of the Multicentre Research Ethics Committee (MREC) at the University of York and the 'York and Humber ethics committee' 11/YH/0028. The project was registered with the NIHR CSP - R&D IRAS project code: 57218.

Initial documentation required modification following the initial ethics review, by the 'York and Humber ethics committee'. The committee considered it potentially too confusing for participants to receive information about the cohort study and the nested treatment RCT together in the information pack and at the same time. Thus a dual consenting process within this study was implemented as an amendment, to fulfil the pilot study recommendation from the 'York and Humber ethics committee' (appendix C1).

The initial consenting process was adapted and separated, and the information sheets and consent forms were redesigned, and two separate consent and information forms were provided for each part of the study. It was decided that distribution of the baseline questionnaire was to occur with the second consent and information sheets. However it was unknown at this stage of the impact, adding an additional layer of documentation to the process would have upon the study. All final study documentation with appropriate version numbers was reviewed and approved by the 'York and Humber ethics committee'.

Samples of the participant recruitment pack were tested internally at the University of York by a number of individuals, prior to the recruitment packs being distributed to participants. Clarity, ease of completion, acceptability to the respondent and the identification of any frequently unanswered questions were assessed in this process and any appropriate adjustments made.

Participants capable of conversing with a physiotherapist in English as detailed in the inclusion criteria were included. As this was a PhD pilot study questionnaires will be available in English versions only. There are financial constraints of a PhD project attached to this study; no provision for translators or additional physiotherapy time that would be required for non-English speaking participants was afforded. For this reason, one of the criteria was a sufficient grasp of the English language. Patients who are ineligible or who are taking part in other research were excluded. An explanation letter was sent to patients who consented but did not meet the inclusion / exclusion criteria.

Once the pre-test of documentation was complete, initially one GP practice was invited to participate and selected as a recruitment and document test site. The rational for this, was to test initial recruitment rates and documentation packs with an initial number of participants, allowing for any modifications to be made to minimal baseline documentation only, if required.

A second letter was mailed to participants, once the initial consent forms (appendix C5) were received by the University. This included a second

information sheet (appendix C6) a consent form (appendix C7) and the baseline (case report form) questionnaires (appendix C8) for those patients who had consented to receive additional information about the nested RCT. Participants not consented to the RCT were sent the baseline questionnaire only.

4.9 General Practice (GP) Recruitment

The GP practices within the York region were identified through an online search of the Primary Care Trust (PCT) and NHS Choices websites in May 2011. A restructure of the NHS in April 2013 saw PCT's replaced by Clinical Commissioning Groups (CCG), which occurred during the period of this study. An aim of identifying ten GP practices was planned, with preliminary investigations into the practice size, locality and involvement within research with the University of York previously being conducted to:

- *Determine the size of the GP practice would assist the trial coordinator in the estimation of how many GP practices would be required*
- *The locality of each practice was important to ascertain for feasibility of participant travel to any potential intervention at the physiotherapy practices*
- *Previous involvement in research with the University of York was important to consider; to establish experience of working with trials but also importantly not to over burden one individual practice with multiple trial involvement requests*

The above information was collected to establish and confirm the suitability of the GP practice for the proposed pilot trial prior to making contact.

Initially five of the ten GP practices were selected based on their locality and travel feasibility for participants to the identified physiotherapy practices. Initially only five practices were invited to participate due to:

- *An unknown response rate of the GP practices*
- *The unknown number of participants to be recruited from each GP practice*
- *The study being a pilot investigation with limited resources*

It was aimed to recruit 2 – 3 GP practices initially. The identified GP practices were invited by letter (appendix C2) to participate within the study. The invitation letter included information regarding the cohort study and nested RCT, and an expression of interest form was included. Interested practices returned their expression of interest forms to the chief investigator at the University of York.

A confirmation letter was then sent to the GP practice (appendix C3) and a discussion and induction meeting was arranged with the GP practices, to gain formal consent and to deliver the participant information packs (containing participant invitation letters, consent forms, a copy of the information sheets, and baseline assessment questionnaires) to the practices. The process of inducting the GP practice was commenced; this involved myself as chief investigator providing a short presentation of the trial, and a question and answer session with key staff members and a discussion session.

At an agreed time, the GP database search was conducted by myself, as chief investigator with a practice representative present to identify a list of eligible patients, and the participant information packs were mailed to all the identified potential participants at the time of the search. The invitation packs were labeled and sent by myself, by post directly from the GP practice, which ensured researchers did not have access to patient data unless a patient expressed an interest in taking part in the study.

Following the successful recruitment of the first GP practice participant recruitment, a second practice was planned for recruitment. If sufficient participants numbers for the purpose of the pilot study were achieved from the two GP practices no further practices were planned to be recruited, if participant recruitment was not sufficient, a further one to eight practices were planned to be invited to join the study. If recruitment exceeded expectation

and participant numbers were higher than required, a two to one allocation method was planned, there by twice the number for participants randomised to the control group than the treatment groups.

Previous trials have shown good recruitment using a GP database recruitment method. A yoga trial for LBP demonstrated a response rate of 12% (994 patients responded from 8,638) (Cox et al. 2010) and an acupuncture trial for Irritable bowel syndrome demonstrated a response rate of 15% (247 responded from 1651) (Reynolds, Bland and MacPherson 2008).

Figure 4.3: Flow Diagram of GP recruitment to participant identification

It was anticipated the recruitment of the required sample size of participants would take between ten and twenty weeks for completion.

4.10 Justification for choosing chartered physiotherapists

Establishing a professional group to take the lead in delivery of prompt and effective care for LBP, is inherently determined by the intervention type, the effectiveness of the professional group and its cost of delivery. I published a paper justifying the use of Chartered Physiotherapists for the delivery acupuncture (Dascanio 2015b), (see appendix A2) which is referenced where appropriate throughout this chapter.

A multidisciplinary approach is often adopted for LBP (Salerno et al. 2002) however professional autonomy within the UK has allowed a more integrative approach with crossover of roles and thus a reduction of the number and cost of staff members. Chartered physiotherapists and other medical professionals

are able to extend their roles (with additional training) and they are able to undertake treatments that would previously have been considered outside their scope of practice (Dascanio 2015b).

Chartered physiotherapists are potentially well placed to help provide a solution and lead on the delivery of interventions for LBP. In the UK there are in excess of 55,000 physiotherapists working across the country, who work at the forefront of healthcare (CSP, 2014a). Chartered physiotherapists are regulated by the Health Care Professions Council (HCPC) and are autonomous holistic practitioners. They have a variety of appropriate skill sets; they have the opportunity to extend their scope of practice and are well established and integrated into the health service with care pathways available to support them. Additionally early referral to physiotherapy for LBP has been shown to reduce the risk of additional healthcare consumption and cost (Fritz et al. 2012).

With reference to the treatment recommendations in the NICE (2009 and 2016) LBP guidelines, many chartered physiotherapists are already specialised and able to deliver the recommended interventions and the NICE recommendations and physiotherapy input are listed below:

Advice on self-care: Physiotherapists routinely offer advice on self-care and self-treatment for LBP, supported by aids such as 'The Back Book' (Burton et al. (2002) and other patient advice material. They also offer advice on posture, correctional movements for activities of daily living and work. Those with additional training are also able to offer ergonomic advice for workstations, seating and car seating (Association of Chartered Physiotherapists in Occupational Health and Ergonomics, 2015).

Exercise (individual & group sessions) – including aerobic activity, movement instruction, muscle strengthening, postural control and stretching: This is a core skill for physiotherapists: they can offer comprehensive exercise programmes on individual and group programmes – incorporating aerobic activity, movement instruction, muscle strengthening, postural control, stretching, Pilates, and yoga. They also often undertake additional

training to advance these skill sets (The Association of Chartered Physiotherapists in Sport and Exercise Medicine, 2017).

Manual Therapy (including spinal mobilisation, spinal manipulation and massage techniques): This is a core skill for physiotherapists; they routinely offer manual therapy – including spinal mobilisation, spinal manipulation and massage techniques. Additional training is encouraged to further develop and specialise in advanced techniques in this area (Musculoskeletal Association of Chartered Physiotherapists, 2017).

Combined physical and psychological treatment programmes: As part of multi-disciplinary teams and with additional training in interventions such as Cognitive Behavioural Therapy (CBT), physiotherapists are involved in treatment programmes and pain management groups nationally, both in the NHS and private sector (Brunner et al. 2013).

Pharmacological therapies: Physiotherapists with additional training have had supplementary prescribing rights since 2005, however an application for full independent prescribing responsibilities was under review for ten years and in July 2012 physiotherapists were awarded full independent prescribing rights. This skill requires additional training, but places physiotherapists well for prescribing medications and providing a complete and comprehensive service for patients with LBP (CSP, 2012).

Acupuncture (Recommended NG88 2009, not recommended NG59 2016): Currently chartered physiotherapists are the largest group of healthcare professionals delivering acupuncture within the UK with over 12% (<6500) of physiotherapists already trained in western medical acupuncture, they deliver acupuncture treatments routinely in the NHS and private sector (Dascanio, 2015a).

Other interventions not currently recommended by NICE included *traction, lumbar supports, injection therapy and electrotherapy modalities* (including laser therapy, interferential, therapeutic ultrasound and transcutaneous electrical nerve stimulation TENS). Though these treatments are not currently nationally recommended they have been historically used for LBP

and can all be provided by chartered physiotherapists if required. They may carry some benefit on an individual case basis, for example, while not routinely recommended using a lumbar support belt for heavy gardening as a preventative measure but not during routine daily life may be appropriate for some individuals. Many of these interventions are often still requested by doctors and supported by some research (Muller, 2009; Watson, 2008; Watson, 2000), though the thresholds of evidence are not sufficient for national guideline recommendation.

Treatment recommendations do change over time with new research published frequently, popular treatments in 2017 may not be considered effective in the future, so it is important health professionals remain current with up to date research and adaptive to change within clinical practice. Many chartered physiotherapists have extended their roles, adapt well to change and would be well placed to deliver interventions for LBP patients. In addition physiotherapists have chartered status, protected titles, have clinical autonomy and are regulated by the HCPC, some professions do not offer this depth of protection to the public. While GPs, nurses, physiotherapists and other healthcare professionals have qualified in the use of acupuncture and other interventions (Dascanio, 2015b), due to their routine knowledge, care and treatment of low back pain and for all the stated reasons, and for clarity and consistency chartered physiotherapists only were used for all intervention delivery for the pilot study for this thesis.

4.11 Physiotherapy practice recruitment

An online search via the Chartered Society of Physiotherapy was conducted of private physiotherapy practices within the York region. It was important the physiotherapy practice would be easily accessible geographically for future participants within the trial from the recruited practices. These practices were selected and contacted by telephone.

An aim of identifying five physiotherapy practices was planned, with preliminary investigations into the practice size, locality and involvement within research with the University of York being conducted:

- *Determining the size of the physiotherapy practice and number of physiotherapy staff available for the study would assist the trial coordinator in the estimation of how many practices would be required*
- *The locality of each practice was important to ascertain for feasibility of participant travel to any potential intervention, considering the recruited GP practice regions*
- *Previous involvement in research with the University of York was important to consider; to establish experience of working with trials but also importantly not to over burden one individual practice with multiple trial involvement requests*

The above information was to be collected to establish and confirm the suitability of the physiotherapy practice for the proposed pilot trial prior to making contact. Private physiotherapy practices only were to be approached for this study, due to the limited time frame for completion of the pilot study and the additional timely layer of ethics involved in recruiting NHS departments and the large department sizes. It would be planned for NHS departments to be incorporated into a full-scale trial.

Initially three of the five practices were selected based on their locality and travel feasibility for participants. Initially only three practices were planned to be invited to participate due to:

- *An unknown response rate of the physiotherapy practices*
- *The unknown number of participants each practice would be able to accommodate*
- *The study being a pilot investigation with limited resources*

The identified physiotherapy practices were invited by telephone and letter to participate within the study. The invitation letter included information

regarding the cohort study and nested RCT, and an expression of interest form. Interested practices returned their expression of interest forms to the chief investigator at the University of York.

A discussion and induction meeting was planned with the physiotherapy practices, to gain formal consent and to provide information and documentation to the practices. It is important that more than one physiotherapist delivered the interventions to avoid any bias of the results being introduced due to a single therapist's expertise.

Recruited physiotherapists were required to be qualified in manual therapy or acupuncture or both. Due to the factorial nature of the nested RCT the same practitioner was required to deliver both interventions in the combined trial arm and for consistency the same type of professional practitioners were required to deliver the interventions to all groups. Chartered physiotherapists were selected for this role due to their ability to deliver both acupuncture and manual therapy (Dascanio, 2015b) justification for this decision is discussed in section 4.10. No traditional acupuncture practitioners were recruited for this pilot.

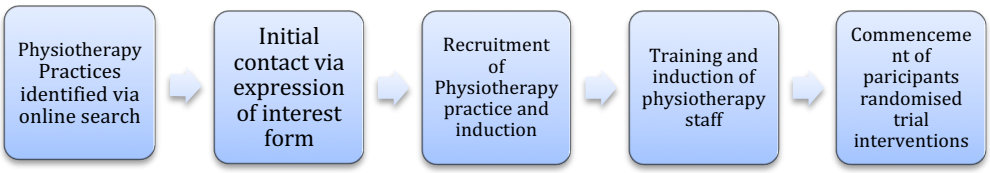

Figure 4.4: Flow Diagram of Physiotherapy practice recruitment process

It was anticipated the recruitment of the required physiotherapists would take between two to six weeks for completion. Following the successful recruitment of the physiotherapy practices, training and induction of all physiotherapy staff was to be conducted at the physiotherapy practices, by myself as trial coordinator.

4.12 Randomisation process

Participants eligible for the nested RCT study were given an identification number. Participant eligibility was determined by:

- *Participants consented to the cohort study*
- *Participants consented to be part of the nested RCT*
- *Participants consented to receive the trial interventions*
- *Participants completed the baseline questionnaire*
- *Participants completed the three monthly questionnaire*
- *Participants who scored 4 or more on the RMDQ*

When a group of participants were found to be eligible for the nested RCT, their identification numbers only were sent to the University of York's Trials Unit Director. The Trials Unit Director acted as the independent data manager and undertook the planned independent random allocation. Randomisation was conducted using the randomization function in SPSS software, such that equal numbers were allocated to the arms within each block. The allocation was not stratified, and the characteristics of the individual participants were unknown to the Director undertaking the allocation procedure (Dascanio et al. 2014). The process was conducted in this way to avoid the possibility of subversion bias.

Randomisation of the participants occurred in a block that was equal to the size of the group. Randomisation occurred when approximately ten participant forms were returned and randomisation was repeated every time a new batch of questionnaires were received. It was anticipated that by randomising in this way, it would allow time for the participants to commence their allocated treatment without delay and not overwhelm the physiotherapy practices with a high volume of participant numbers simultaneously (Dascanio et al. 2014).

The size of the control group depended upon the numbers recruited to and from the cohort, but it was to be at least as large as the intervention groups. It was planned that if recruitment to the cohort and thus the RCT was greater

than anticipated, the usual care group would be larger than the treatment groups; to enable an unequal allocation ratio favoring the control group compared with the intervention groups. This would afford the trial to have greater power through an increased sample size and be more efficient. The number of participants recruited determined the allocation ratio. If recruitment meant that more patients were available for the study than could be treated in the intervention groups, then the rationale of putting more into the control group means that the total sample size can be increased to improve the power of the study.

Prior to randomisation participants were asked which treatments they would consider for their low back pain. If a participant expressed an unwillingness to receive one or all of the specific treatments (i.e. manual therapy or acupuncture), they would still be included in the study, but randomised to usual care or the remaining intervention that they did not express an aversion to. Participants who did not wish to receive any intervention would be followed up in the observational cohort study and would not be randomised.

4.13 Methods for preventing bias

Randomisation within the RCT, if conducted effectively, will eliminate selection bias and an intention to treat analysis principle was used as it is the most robust analysis technique for preventing the introduction of bias (Torgerson and Torgerson 2008).

As discussed in chapter three, the majority of attrition occurs within the first follow-up period of an RCT and is a major threat to internal validity of a study (Relton et al. 2010). Using a cohort design with nested RCT will allow a 'run-in period' to occur, allowing for the collection of baseline data and a set of three-month outcome data to be collected, prior to randomisation. Only participants returning their three-month questionnaire were eligible for randomisation. Therefore subsequent attrition post randomisation was expected to be minimal. Further attempts to reduce attrition were planned through the provision of comprehensive explanation of the study and regular

contact in the form of questionnaires to maintain regular contact with the participants throughout the study (Dascanio, 2015d).

Other potential sources of bias are dilution effects (i.e. some participants, randomised to intervention fail to accept the allocated treatment) and patient preference, however it is anticipated this will be low due to nature of the cohort recruitment design methods in offering an element of choice in the early phase of recruitment (Brewin et al. 1989). As participants were entering a cohort and aware that they may be approached to take part in a trial, but aware that only a small portion of participants may be offered treatment, informative explanations were provided to the participants (Torgerson, Kaber-Moffett and Russell, 1996).

As this was a pragmatic study, it was not possible to blind the participants or the clinicians to the allocated interventions. Those allocated to the usual care group however were not advised of their allocation, only that they were continuing within the cohort, thus limiting resentful demoralisation within this group. Both the participant and their GP were advised if they were assigned to one of the treatment groups. As the final outcome assessment was provided by the participant no blinding of the outcome measurement occurred (Dascanio, 2015d).

4.14 Duration of treatment period

The treatment period for the trial interventions was between ten and twelve weeks. The allocated treatment intervention aimed to be conducted once per week for a period of ten weeks; however a two week threshold to twelve weeks allowed the ten allocated treatment sessions to be completed if any delayed or missed treatment sessions occurred due to sickness or other reason. This timescale was in line with the recommendations from the NICE guidelines for LBP (NICE, 2009).

4.15 Outcome measurement and follow-up

On entering the study each participant completed a baseline assessment delivered via a postal questionnaire, which included RMDQ and MODI, (detailed in section 4.5). This was then followed up by a postal questionnaire at three months. On completion of the three-month questionnaires, eligible and consented participants were randomised to one of the four trial arms.

Follow-up then occurred at six months via postal questionnaire; this timing coincided with the completion of treatment for the active treatment intervention groups. Further follow up via postal questionnaires occurred at, 9, 12, 15 and 18 months, however this thesis only considered the data to the six-month follow up period, figure 4.5 demonstrates the flow diagram of participants through the study.

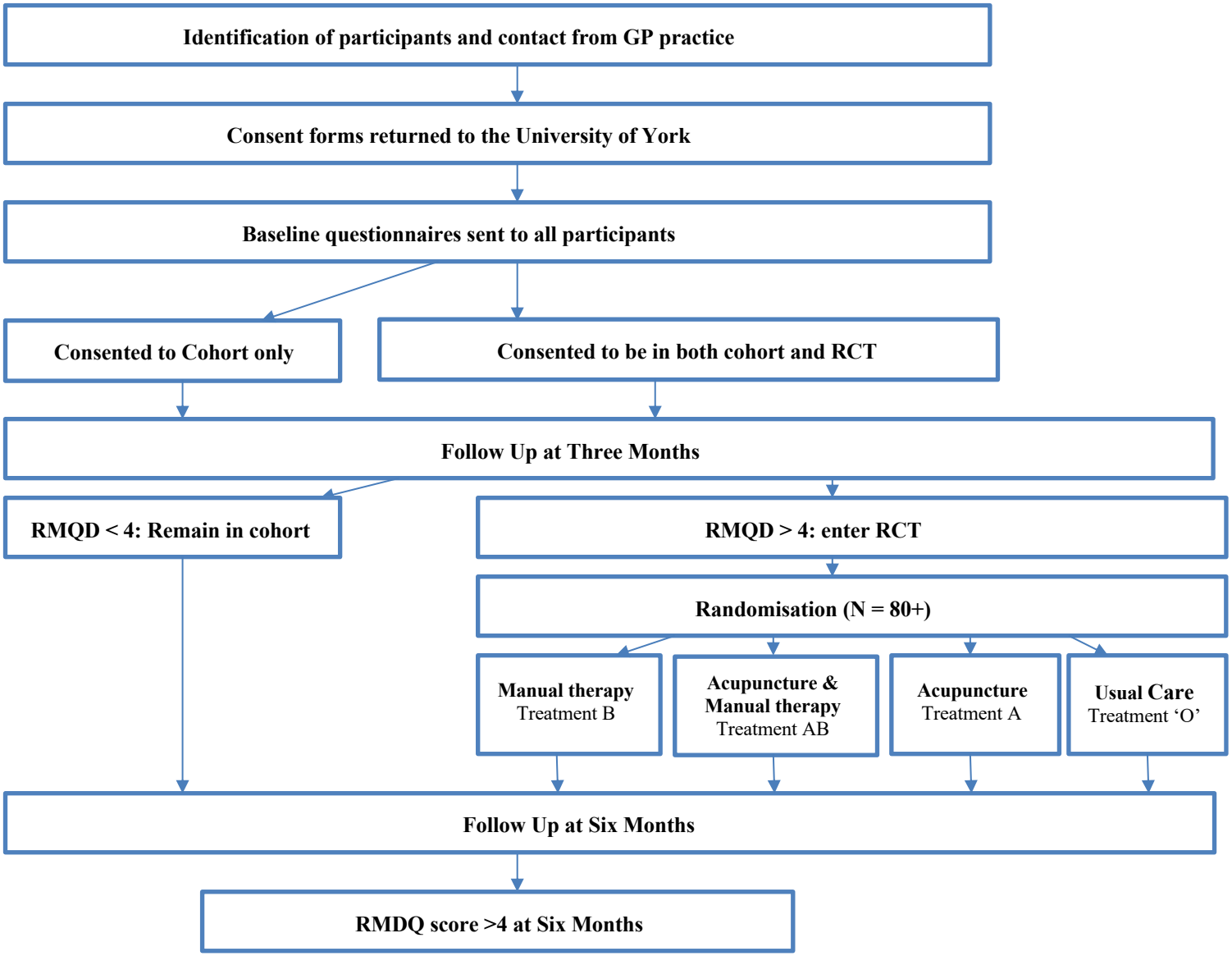

Figure 4.5: Flow diagram of participants through the study (Adapted from Dascanio, 2015d)

4.16 Data Collection

Participant baseline data was collected and inputted into an excel spreadsheet by the trial coordinator; each participant was issued with a trial participant ID number. When the baseline, three and six questionnaires were received, the receipt of these was recorded onto an excel spreadsheet using the participant ID numbers to ensure all appropriate questionnaires had been received.

All baseline and follow-up data were collected on paper questionnaires called Case Report Forms (CRFs) and identified solely by the trial Participant ID. The CRFs were returned by post to the University of York where they were scanned, using Teleform data capture software, into a bespoke data management system. The data were error checked and validated to ensure accuracy. The paper CRFs were held securely in a controlled access area in locked cabinets but separate from the consent forms.

4.17 Missing Data Follow-Up

If postal questionnaires were not returned at three months or subsequently, it was planned to send an initial text reminder to the participant at ten days, followed by a reminder telephone call at two weeks and a further follow-up telephone call was also planned at four-weeks.

If missing data were evident on the case report form questionnaires, participants would be contacted by telephone in an attempt to retrieve the data, by asking the questions to the participant over the telephone in order to support the participants if they found difficulty completing the questionnaire. A second follow-up call was planned if the first contact attempt was unsuccessful. In a deviation from this protocol the process for collecting missing data was only applied for the RMDQ and the MODI, as these data were analysed immediately following the completion of the pilot

study. No attempt was made to collect missing data from all the ancillary data (SF-12, EQ-5D-3L, VAS, Health Utilisation Questionnaire) due to the missing data for these data sets only being reviewed when the statistical analysis was conducted a considerable time (several years) after the pilot study was completed.

4.18 Sample size calculation

This pilot study was not planned to determine or compare effectiveness between groups, and not powered to do so, as is common practice for pilot studies (Lancaster et al. 2004). Sample size estimation for a pilot study is not straightforward, as the outcomes of a pilot study tend to be process measures rather than quality of life or clinical outcome measures. Though this study will collect outcome data to inform a full-scale study.

Potential recruitment of 16 participants for each intervention group would be acceptable; as this exceeds the minimum recommended number of 12 per group for a pilot trial, as is routinely recommended (Whitehead et al. 2016; Moore et al. 2011; Julious 2005) and if aiming to power the study, would enable the study to have an 80% power with a significance level of 0.05 to observe a 1 standard deviation difference between the two groups, if a difference were to exist (Bland, 2000).

In the proposed pilot study no formal sample size calculation was conducted. However it was planned to recruit a sample size of a minimum of 16 participants per group. Allowing for any attrition and the pragmatic nature of the trial, recruiting 16 - 20 participants to each arm of the trial, therefore a sample size of 64 - 80 will be a conservative target and achievable within the limitations of the pilot (Whitehead et al. 2016; Moore et al. 2011).

While the objective was not to consider effectiveness, the target sample size planned would allow an observation of 1 standard deviation difference between the usual care and treatment groups, powered to 80% (Dascanio,

2015d) this would assist in determining the data requirement for a sample size calculation to inform a full-scale study and also to mirror a planned full-scale study as closely as possible.

An anticipated recruitment rate was not established prior to the study. The actual study recruitment rate would be used to provide information to establish an anticipated recruitment rate for the planning of a full-scale study.

4.19 Anticipated problems with treatment compliance

A treatment compliance rate of 90% was anticipated for the interventions within this pilot study. Evidence from the UKBEAM trial demonstrated compliance of 92% for the manipulation intervention group (UKBEAM, 2004) and a similar compliance rate of 91% was seen for the acupuncture intervention group in a trial of LBP (Leibing et al. 2002). Participants within this study attendance was planned for one weekly treatment sessions and in the combined treatment group the treatment occurred at the same time, a similar compliance rate for this study was expected.

4.20 Minimising loss to follow up

A 20% loss to follow up rate was anticipated for this study as is in line with other similar studies. There was 24% attrition in an acupuncture trial of LBP (Leibing et al. 2002) trial over a nine-month follow-up period and 33% was lost in the UKBEAM trial (2004) over a twelve-month period.

However the cohort study design with a nested pragmatic factorial RCT has been selected for this pilot to consider and review its effectiveness in minimizing trial attrition, therefore a lower rate of attrition was potentially expected.

If a participant wished to withdraw from their intervention at any time they were asked if they were happy to complete the appropriate questionnaires

for our analysis using the intention to treat principle, the participants were free to accept or decline this request.

4.21 Methods of analysis

The CONSORT guidelines for randomised control trials will be used for the reporting of this study and data analysis (Moher et al. 2001). An intention to treat (ITT) principle will be applied: therefore all participants will be included and analysed in their original randomised groups, regardless of completion of their intended intervention (Torgerson and Torgerson 2008).

Data will be collected at three monthly intervals, with a single principal analysis occurring at six months, following the trial interventions.

4.21.1 Primary analyses

The results for this pilot study are largely descriptive and reflect upon the study design for implementation of a full-scale trial and they focus upon:

- *An analysis of the recruitment rates of GP practices to the study*
- *An analysis of the recruitment of physiotherapy practices and practitioners to the study*
- *An analysis the number of participants recruited to the cohort study from each GP practice*
- *An analysis of recruitment rates of participants from the cohort study to the nested factorial RCT*
- *An analysis of the participant consent rates and acceptance of the study interventions*
- *An analysis of the success of delivering combined interventions by the practitioners*
- *An analysis of the attrition rates of participants, pre and post randomisation within the cohort and nested RCT*
- *A review of baseline characteristics of the study participants*

- *A descriptive analysis of the use of the design methodology for evaluating individuals suffering with LBP*
- *A sample size calculation for a full scale study*

4.21.2 Sub-analyses

The planned preliminary investigation involved estimating the effect of:

- *Acupuncture alone versus usual care*
- *Manual therapy alone versus usual care*
- *Combined acupuncture and manual therapy versus usual care*
- *Combined intervention compared to each of the single treatments*

The estimation of effect was considered for both the RMDQ score and MODI score at six months (three months post randomisation) to determine the standard deviations for calculating a sample size, to enable the comparison of the outcome measures and to observe if any directions of effect or associations were evident to inform a full-scale study.

For each comparison, analysis of covariance was employed adjusting for the score reported immediately prior to randomisation (at three months) (hereafter referred to as the "screening score") to obtain treatment estimates with 95% confidence intervals. The trial was not powered to detect a specific difference between intervention groups and all analyses therefore were exploratory only to inform a full-scale study. Participants identified who were unwilling to accept randomisation to all interventions were excluded from any comparison analysis between acupuncture and manual therapy (alone or in combination).

Continuous data are to be summarised as mean and standard deviation (SD) and categorical data as frequency (%), with adjustment for screening scores.

Sub-analysis included:

- *An analysis of the use and completion rates of Roland Morris disability questionnaire (RMDQ) and Modified Oswestry Disability Index (MODI)*
- *A regression analysis of mean scores will be performed adjusting for baseline assessment for the Roland Morris Disability Questionnaire – (an estimate of standard deviations will be calculated and combined with acceptance and attrition rates for a future sample size calculation)*
- *A regression analysis of mean scores will be performed adjusting for baseline assessment for the Modified Oswestry Disability Index*
- *A regression analysis will compare the three treatment groups to the control group. Analysis of differences between the combined treatment group compared to the single treatment groups. The study is of relatively low power to demonstrate any difference, any results will be treated with caution and are primarily intended to inform the design of a future study*
- *Comparison of the characteristics of the two outcome measure tools (RMQD and MODI) to assess which would be most appropriate for a full-scale trial*

4.22 Economic data

The NICE guidelines (2009) calculated the additional cost to the nation of implementing acupuncture for the treatment of LBP would be £24,366,000, however this estimation failed to consider the use of services already in place in the NHS, for example within physiotherapy departments, and the estimation was based on the introduction of new staff and services. The potential additive benefits of combining two treatments, was also not considered by NICE (2009; 2016).

This pilot study will aim to inform a full-scale trial if the incorporation of acupuncture into an already existing physiotherapy session is a viable option.

4.23 Ethics and governance

Prior to the commencement of the trial, ethics approval and research governance was sought from the MREC (Multicentre Research Ethics Committee) at the University of York Health Sciences Committee and the York and Humber ethics committee. The study commenced once all approvals were in place as detailed in section 4.8.

4.24 Data protection and confidentiality

All participant documentation is kept in line with the Data Protection Act (1998) and paper copies will be retained for 7 years after the completion of the trial. All electronic data was password protected on secure computers.

All personal information remained confidential and was anonymised through the use of coding as appropriate.

4.25 Monitoring of adverse events and safety protocol

A data and safety monitoring committee was formed to monitor the trial and any adverse effects that occurred. The committee included myself as trial coordinator, the trials unit director (DT) and another trials unit staff member (LC). A safety reporting protocol was established in accordance The York Trials Unit, Standard Operating Procedures (SOP) (version 2.0) and with the NIHR clinical trials tools kit (NIHR, 2017b), and followed in the event of any safety concerns or adverse events (appendix C12).

The treating chartered physiotherapists were professionally responsible, trained and competent to determine the level of a safety issue or adverse event, if any occurred and thus make a clinical judgement for the appropriate procedure, as expected within their professional code of conduct. The trial coordinator would be informed of any incident and a

report made to the data and safety monitoring committee, for any actions or recommendations to be made.

Acupuncture and manual therapy are rarely reported to cause adverse effects (White, 2006; White et al. 2001: MacPherson et al. 2001). However, adverse events or safety issues can happen to participants and practitioners in trials. During the trial, if any occurred to the participants, the physiotherapist assessed them for the seriousness and the nature of the incident i.e. was it related to the intervention and was it serious, and then the individual was treated locally or referred to the appropriate location, i.e. A&E, local hospital or GP, as per the seriousness of the incident. If a physiotherapist experienced an adverse event or safety issue i.e. a needle stick injury, physiotherapists were required to assess themselves, to determine the seriousness and nature of the incident.

The seriousness of a reaction determined the action taken e.g. a needle stick injury would require attendance at A&E, a participant experiencing considerable pain following an intervention would require a GP referral. It was planned for the trial coordinator to follow up any individuals who experienced adverse events or safety concerns, and detailed information and any treatment delivered was documented and reported to the safety monitoring committee.

In the event of minor reactions, such as feeling faint (commonly experienced with acupuncture) participants would be treated locally as appropriate by the treating therapist until symptoms resolved and the GP and the data and safety monitoring committee would be kept informed of any concerns (Dascanio, 2015d).

All centres providing acupuncture were registered and approved by their local health authority. Sharps policies and Health and safety polices were in place. A needle stick injury protocol was also in operation (MacPherson et al. 2004).

4.26 Informed consent

Participation in the study was entirely voluntary. Written informed consent was required from all participants. Entry into the cohort or RCT was not permitted without informed consent.

A clear and easily understandable participant information sheet (appendix C4) was provided to all consented potential participants. It stated not entering the trial, withdrawal of consent or withdrawal from the trial at any time in no way affected their present or future quality of care. It aimed to provide an unbiased explanation of the treatment interventions investigated, including any potential benefits and known risks (Dascanio, 2015d).

4.27 Dissemination of findings

Debriefing and sharing of the study results and findings with study participants who wished to know the outcome of the study is standard practice via the University of York website. However the Research Government Framework published by the Department of Health (DoH, 2001) outlined that sharing of the study outcomes and also information regarding a participants' individual treatment allocation and their outcome should be standard procedure in clinical research. Further consideration and input in this matter would be required for a full-scale RCT, as individual participant data was not available to distribute from this pilot study.

Dissemination of this study aimed to be extensive to help raise awareness of the potential need for research in this area. Findings of the study were presented through peer-reviewed publications and conferences (local, national and international). A single piece of research may not in itself change practice however it aimed to initiate discussion and therefore interest in the area, potentially attracting future funding for a full-scale trial

that may provide clarity of the NICE guidelines and pathways for LBP (Dascanio et al. 2014).

It is intended for the findings of the study to be presented in high impact peer-reviewed publications, which are accessed as many health professionals and policy makers as possible. The results of the study will be submitted for presentation, as an abstract or poster presentation at conferences working in the field of LBP, acupuncture and manual therapy in the UK and internationally as appropriate (Dascanio, 2015b, c, d). All active contributors will be credited in the main report and publications.

4.28 Discussion

The planned study aimed to pilot a relatively new design methodology, which had not previously been used for the musculoskeletal disorder of LBP. The study considered the role of a cohort design to aid recruitment to a nested RCT and evaluate the amenability of participants following the methodology.

The key objectives of the pilot study were to investigate the practicality of it being run as a full-scale trial, to pilot the use of the cohort methods to recruit and retain participants to the trial, to establish appropriate outcome measures, review recruitment and retention, to investigate the value and functioning of the factorial methodology, and to assess the achievability of combining two treatments within a single treatment session. A factorial design was used to consider the efficiency of running two trials within one.

It is the intention of a full-scale trial to maximise generalizability of the study while maintaining strength within its internal validity and contributing to the research base for care pathways for LBP. The pilot study aimed to aid future decisions for a full-scale trial with information and knowledge gained from the running of the pilot study.

The strengths and weaknesses of this study will be discussed in the final chapter of the thesis. The following chapter will present the results of the conducted cohort study with nested factorial RCT.

5 Results of the Pilot Study

5.1 Introduction

In chapter two, I discussed the need for more evidence to ascertain the effectiveness of acupuncture in order to reduce uncertainty and inform guideline developers and clinical practice. In chapter three I demonstrated why a cohort study with nested factorial RCT had features that would provide robust evidence in an efficient way for a study of acupuncture and manual therapy for LBP. In chapter four I presented the methods of the planned pilot study.

In this chapter, I report the results of the pilot study and critically examine the implementation of the processes of the pilot study. I will report: recruitment rates of participants to a cohort study from GP practices, and of participants from the cohort study to a nested RCT, rates of consent to participate and acceptance of the interventions by participants and delivery by therapists, and attrition rates of participants pre and post randomisation. I also report on the use of factorial RCT design for a population with low back pain. I will report on the use of the outcome measurement tools, which will include an updated exploratory analysis of the outcome tools (RMDQ and MODI) used within the study and these will be presented and discussed to inform a full-scale trial. I will also present the results from a range of outcome measures: the SF12, the EQ-5D-3L, the VAS pain scale and the health utilisation questionnaire.

Some of the results for this pilot study have been published (Dascanio 2014, appendix A6 and updated in A6b), and is referenced where appropriate throughout this chapter.

5.2 Recruitment

5.2.1 General Practice (GP) recruitment

A total of ten potential GP practices were identified within the York area as presented in chapter four. An initial five practices were selected from the ten GP surgeries identified, based on their locality and travel feasibility for participants to the identified physiotherapy practice.

GP Practice one: GP practice one was based York and had a practice list of 17,564 patients. A meeting was arranged with the first GP practice Data Quality Manager on the 27th July 2011 and the practice was successfully recruited. The GP practice was inducted, the database recruitment process was established and a timescale for the start of participant recruitment was identified and agreed.

The GP practice operated the 'Synergy database' system to store patient information. Although older than other databases, the synergy database system was comparable to that used in other practices and permitted group searches of participants to be conducted. The patient population was 17,564 at the time of recruitment. 275 (1.6%) of the total patients were identified as eligible for the study and invited by letter (appendix C4) to participate. 25 (9%) of the invited participants were recruited from the first practice over a three to twelve week period, as shown in figure 5.1.

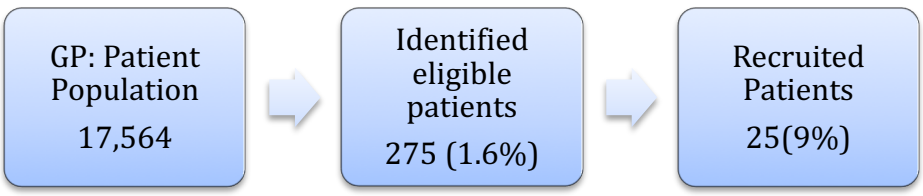

Figure 5.1: Patient recruitment from GP Practice one

GP Practice Two: GP practice two was based in York and had a practice list of 13,538 patients. The GP practice experienced a change in personnel during the preliminary discussions and after a period of uncertainty about

their continued involvement in the trial, a meeting was arranged with the Research Liaison Manager at the practice on the 18th October 2011 and the practice were recruited. The same induction process for practice one was used for practice two and a timescale for the start of participant recruitment was identified and recruitment commenced.

The second practice operated the EMIS LV clinical system software. The patient population was 13,538 at the time of recruitment. 570 (4%) of all patients were identified as eligible for the study and invited by letter to participate in the study. 100 participants (18%) were recruited from the second practice over a shorter time scale of a one to six week period, as shown in figure: 5.2.

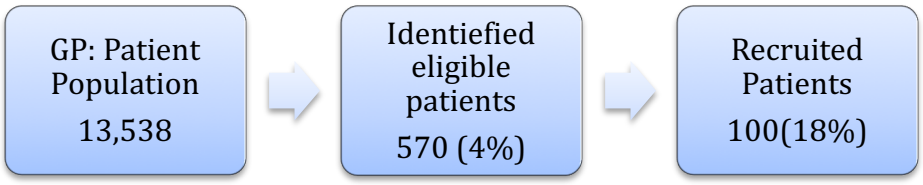

Figure 5.2: Patient recruitment from GP Practice two

In figure 5.3 the GP practice recruitment process for this study is illustrated. Following the successful recruitment of 125 participants from the two initial GP practices and a limited time frame for recruitment, it was decided not to expand the invitation to any further GP practices.

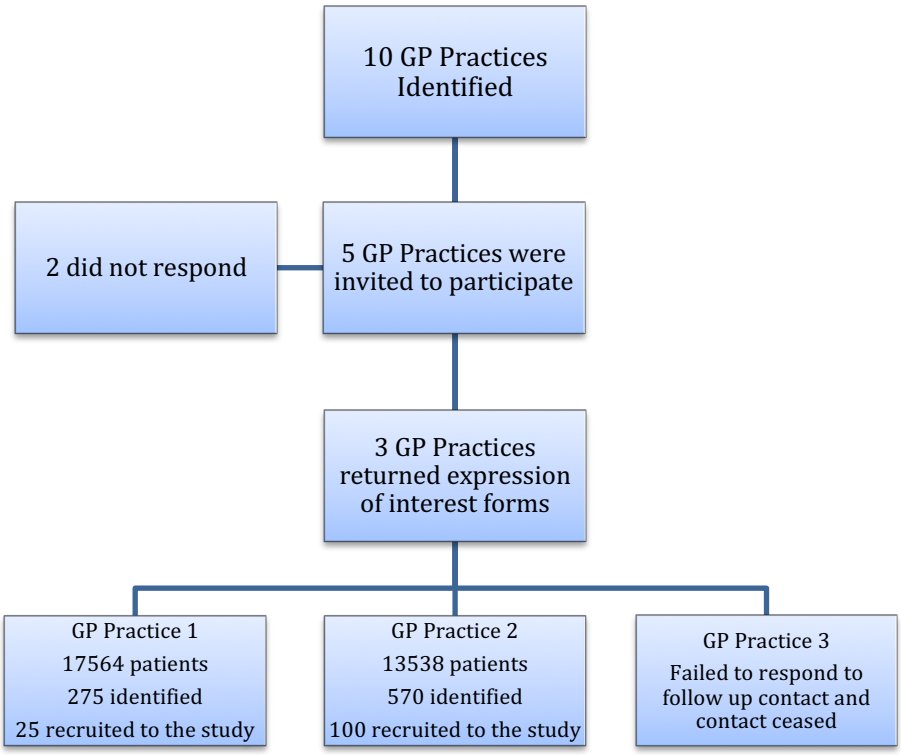

Figure 5.3: GP practice recruitment flow diagram

5.2.2 Physiotherapy practice recruitment

Five private physiotherapy practices were identified. Three of the five physiotherapy practices were local to the area of the University of York. In figure 5.4 the recruitment process of physiotherapy practices is illustrated.

Private physiotherapy practices were approached to participate in this study, due to the limited time frame for completion of the pilot trial. Of the three invited physiotherapy practices, one consented to participate, one declined to participate after initial interest, and one had recently closed down for business.

The participating practice employed eight physiotherapy staff, five were qualified to practice acupuncture, and sufficient staffing was therefore considered to be in place to conduct the study at a single practice. By the start of the RCT recruitment however, the physiotherapy staffing level had fallen from eight to three members of staff, two of whom were acupuncture

trained. While the practice were still able to participate in the study their staff numbers had significantly reduced, and time did not permit the recruitment of any further physiotherapy practices to the pilot.

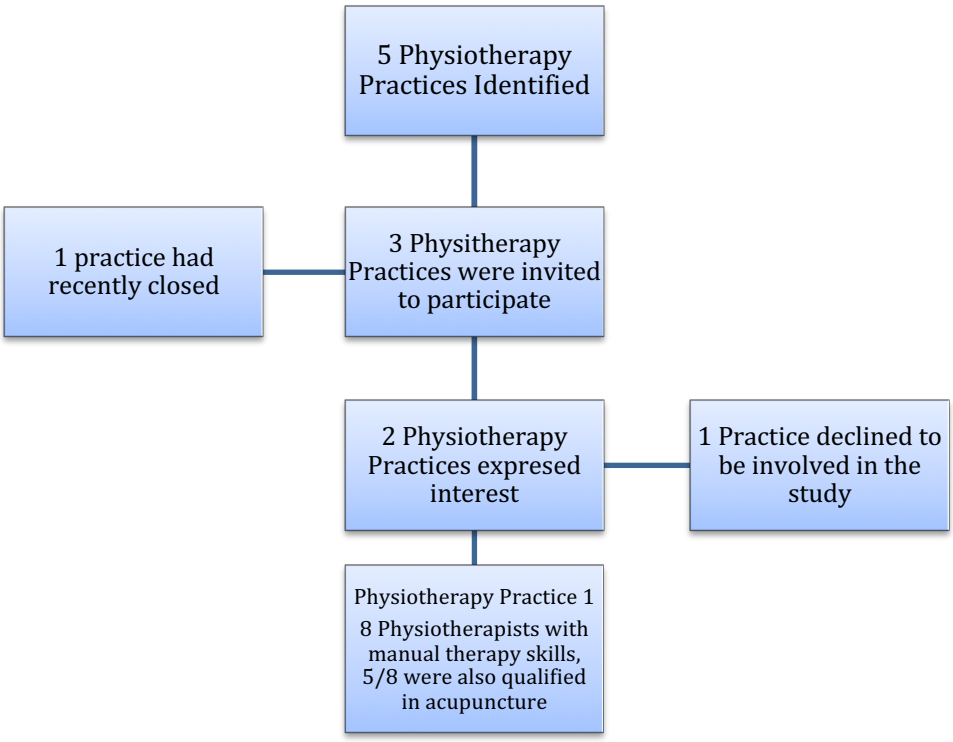

Figure 5.4: Physiotherapy practice recruitment flow diagram

5.2.3 Participant recruitment

Recruitment of participants to the study occurred between July and December 2011 from the two recruited GP practices. Participants who had reported to their GP with LBP within the preceding 18 months were identified on each GP database system; the exclusion criteria were implemented during the selection process to provide potentially eligible participants, as displayed within table 5.1. The details of all the exclusion criteria and their rationale are detailed in table 4.3 (chapter four).

	GP Practice 1	GP Practice 2
Number of patients at GP practice prior to exclusions	17564	13538
Reporting LBP & between the ages of 18 to 65 years, excluding pregnant ladies	293 (1.7%)	757 (5.6%)
After excluding:	-	-
- Rheumatoid arthritis, ankylosing spondylitis, immune compromised, anti-coagulants, heparin, anti-platelets	280 (1.6%)	642 (4.7%)
- Metal allergy or needle phobia	279 (1.6%)	639 (4.7%)
- With psychosis, bi-polar, drug and alcohol abuse	277 (1.6%)	577 (4.3%)
- Valvular heart disease, demand pacemakers	275 (1.6%)	576 (4.3%)
- Neurological conditions, MS, Parkinson's, meningioma	(excluded in group searching system prior to 293 start figure)	574 (4.2%)
- Those unable to consent	(excluded in group searching system prior to 293 start figure)	570 (4.2%)
Total eligible following exclusions	275 (1.6%)	570 (4.2%)

Table 5.1: Results from GP database search, applying exclusion criteria

A total of 845 (2.7%) of potential participants were identified. I sent a letter as the lead researcher to all 845 patients from the GP practice on the same day as the searches were conducted.

In figure 5.5 below the CONSORT flow diagram of participants through the study from invitation to the six-month time point is illustrated, further discussion also follows in the narrative.

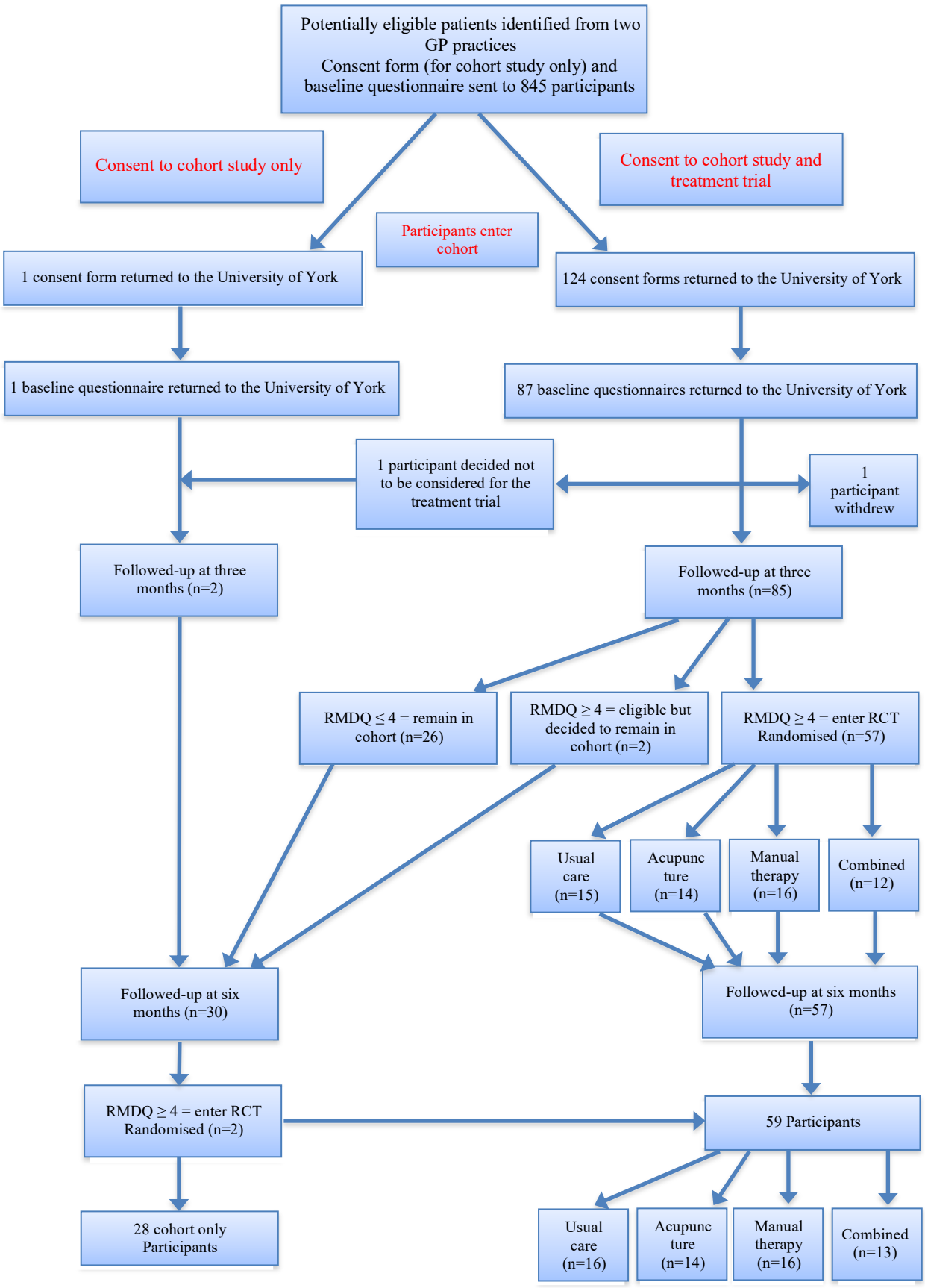

Figure 5.5: CONSORT flow diagram (adapted from Dascanio et al. 2014)

5.2.4 Recruitment rates to the cohort from GP practices

Participants consenting:

125 consent forms were returned between August and November 2011, from the 845 patients invited. Yielding a response rate of 15% via a GP database recruitment process.

Of the 125 patients, one patient consented to enter the cohort study only, and did not wish to be considered for the treatment trial. 124 patients consented both to participate in the cohort study and to receive further information about the treatment trial.

Participants returning baseline questionnaires:

88 of the 125 consented patients returned their baseline questionnaire; a response rate of 70%. Consented patients who did not return their baseline questionnaire were not followed up and not entered into the cohort study. Information regarding any differences between those who did and did not return their questionnaires was not investigated in this study; this is discussed further in chapter eight (section 8.3).

Of the 88 patients who returned their baseline questionnaires and trial consent forms, one additional patient decided not to be considered for the treatment trial and choose only to continue in the cohort study. Therefore two participants were eligible for the cohort study only and a further 86 were eligible for the cohort study and available for the nested RCT. Four weeks after their consent, but prior to the three monthly questionnaires being distributed one additional participant decided to withdraw from study completely, reporting they were pain free.

Participants returning three monthly questionnaires:

At the three-month time point the 87 participants were sent a questionnaire. In my nested study design this is the time point at which the cohort participants became eligible to enter the RCT, thus providing the run-in period within the recruitment process. All 87 participants, (100%) returned their questionnaires fully completed.

Of the 85 potentially eligible participants who had consented to be part of the nested RCT, a further two opted to remain only within the cohort study prior to randomisation, therefore four participants in total remained in the cohort only.

26 other participants scored less than four on the RMDQ and were ineligible for the nested trial so continued in the cohort study only. This corresponded to a 30% loss of potential participants from the cohort population.

30 participants continued in the cohort study and continued to complete their questionnaires; only one participant contacted the investigator to enquire when they would be entered into the treatment study. Involvement in a cohort study remained acceptable to this group of participants reporting low levels LBP on the RMDQ.

5.2.5 Recruitment rates from the cohort to the nested RCT

Participants entering the nested RCT:

The remaining 57 participants scored four or above on the RMDQ and were subsequently recruited to the treatment trial. The participants were allocated randomly to one of four groups using the randomisation function within SPSS (detailed in chapter four).

At the six-month time point, 11 of the 26 cohort participants who were previously ineligible to participate in the trial (because of their low RMDQ scores) returned their questionnaires with a score of 4 or more on the

RMDQ. The increase in their RMDQ score rendered them now eligible for participation and randomisation into the nested RCT.

Two of these 11 eligible participants were invited and entered the trial at the six-month stage. Nine were not invited to enter the treatment trial due to their six-month time point occurring too late in the process and therefore they remained within the cohort study.

A total of 59 participants, therefore, were randomised into the treatment trial and remained in the trial, (57 participants were recruited from the cohort into the nested RCT at three months and two entered at the six month time period).

16 participants were allocated to usual care, 14 participants were allocated to the acupuncture arm, 16 participants were allocated to the manual therapy arm and 13 participants were allocated to the combined acupuncture and manual therapy arm as shown in table 5.2.

The minimum recruitment of 12 participants per group recommended for a pilot trial (Julious, 2005) was achieved for this pilot. However, the study's sample size target of 16-20 participants per group (total sample size 64-80) was not achieved.

Acupuncture	**Manual Therapy**	
	B (Yes)	O (No)
A (Yes)	13 Acupuncture and manual therapy (cell AB)	14 Acupuncture (cell AO)
O (No)	16 Manual therapy (cell BO)	16 Usual GP Care (cell OO)

Table 5.2: Participant allocation to trial intervention groups

Summary of participant recruitment:

- *An initial patient pool of 31,102 were registered at two GP practices*
- *845 patients were identified with LBP and invited to participate*
- *The initial response rate of returned consent forms was 15% (125/845)*

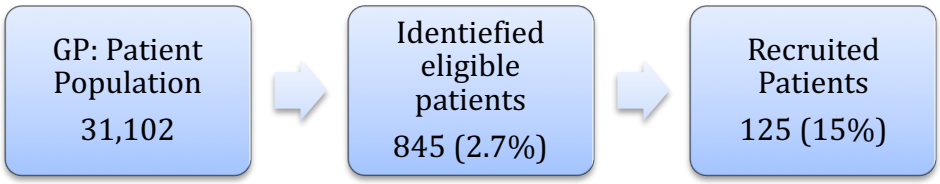

Figure 5.6: GP participant recruitment (pooled data)

- *70% (88/125) of consented patients returned their baseline questionnaire and entered the cohort*
- *Of the initial 845 patients, 10% (88/845) entered the cohort study*

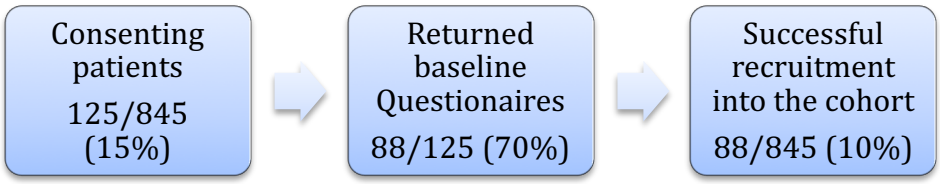

Figure 5.7: Participant recruitment to the cohort study

- *Recruitment from the cohort study to the nested pilot treatment trial yielded a 65% (57/88) recruitment rate at the three month time point.*
- *Two participants requested to remain in the cohort only*
- *One participant withdrew from the cohort study (prior to randomisation)*
- *30 participants were recruited into the cohort study only, at the six months, none of the 30 participants had withdrawn*
- *26 previously ineligible participants became eligible to recruit from the cohort to the RCT at the six month time point*

- *Two of the 26 consented participants entered the trial at six months*
- *28 participants remained in the cohort only at the conclusion of this phase of the study*
- *Of the initial 845 patients mailed from GP practices, 7% (57/845) entered the nested RCT via the cohort study*

5.2.6 Consent and acceptance of trial interventions

97% (82/85) of cohort participants consented to the nested trial and to all the treatment interventions. A high acceptance and consent rate to the trial interventions: acupuncture and manual therapy was observed.

Of the 85 participants, two were unwilling to receive acupuncture and so they were randomised to either usual care or manual therapy only. One participant was unwilling to receive manual therapy and they were randomised to usual care or acupuncture only. These participants self-excluded from acupuncture or manual therapy, they were able to be randomised to the available two groups and remain within the trial. The three participants were not included in any statistical comparisons conducted between acupuncture and manual therapy, alone or in combination.

All participants allocated to an intervention treatment group completed their course of ten treatment sessions, one participant required 13 weeks to complete his 10 sessions due to a holiday and sickness, but was able to complete the course with additional time. Thus there was a 100% attendance rate, demonstrating a very high level of compliance and treatment acceptance.

The chartered physiotherapists were able to deliver all the interventions and successfully combined acupuncture with manual therapy within a single treatment session for all the required sessions.

5.2.7 Attrition rates of participants pre and post randomisation

Figure 5.8 illustrates the flow and attrition rate of participants throughout the study from consent to the six-month time point.

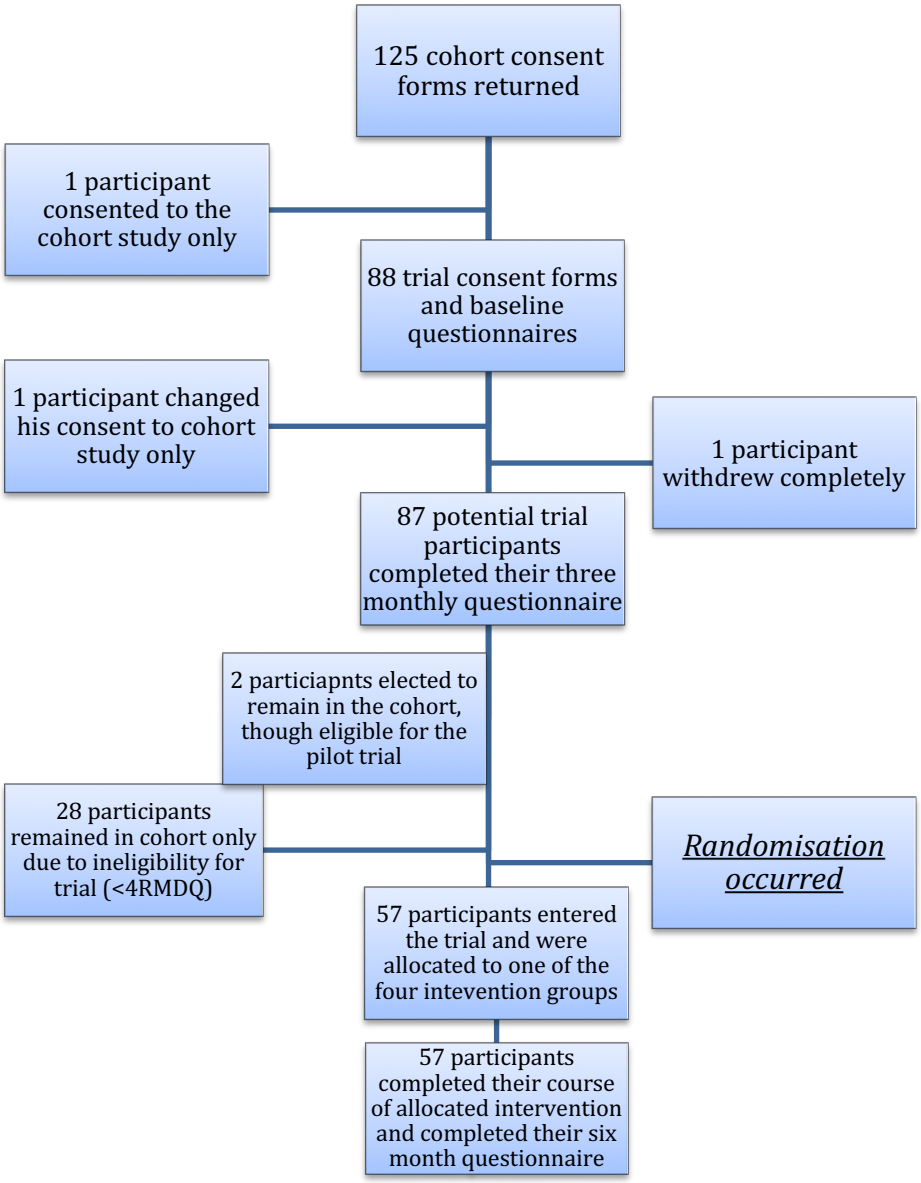

Figure 5.8: Attrition throughout the study, flow diagram

Prior to randomisation 70% of consented patients returned their baseline questionnaires, a loss of 37 potential participants. Attrition from the baseline questionnaire to the time point just prior to randomisation (three-

month time point) was low at 1%, with one participant withdrawing completely from the study (87 participants remained in the study). One additional participant elected to remain in the cohort only; both decisions were made prior to randomisation.

Of the 30 participants eligible for the cohort only, all 30 remained at the six-month time point. Zero attrition for the cohort between the three and six month time periods occurred.

Of the 57 participants who were randomised into the treatment trial at three months none withdrew from the trial at the six-month follow-up time point, post randomisation (95% CI for attrition 0.0% to 6.3) (Dascanio et al. 2014). Providing a zero post randomisation attrition rate for this nested RCT design.

5.3 Baseline characteristics

Baseline characteristics are described at the time point for entering the nested RCT, at three months following the run in period within the cohort study.

5.3.1 Baseline characteristics of the participants

As shown in table 5.3, screening demonstrated the mean (SD) age of participants of the cohort study was 46.3 (9.6) years. At randomisation for the trial participants it was also 46 (12) years, with a range 19-64 years.

Any differences occurring between the intervention groups in the trial should be random as they were allocated randomly and reflects the small numbers involved, but it is of interest to view these differences at baseline to observe any potential anomalies and identify if future stratification at randomisation may be required.

Characteristic	Cohort study only (N=30)	Nested RCT only (N=57)	Whole study (N=87)
Age, years Mean (SD)	46.3 (9.6)	46.5 (12)	46.4 (13.7)
Male (%)	8 (29%)	23 (39%)	32 (37%)

Table 5.3: Age and Sex demographics at baseline, of cohort only, nested RCT only and whole study

As shown in table 5.4 participants in the combined intervention group are observed to have a mean age approximately 5 years older than patients in the cohort and other trial arms. Participants in the manual therapy group were slightly younger (<2years) compared to the other groups.

Characteristic	Cohort study (N=30)	Usual care (N=16)	Acupuncture (N=14)	Manual therapy (N=16)	Combined (N=13)
Age, years Mean (SD)	46.3 (9.6)	46.3 (11.3)	45.6 (11.9)	43.9 (13.7)	50.1 (9.3)
Male (%)	8 (29%)	5 (31%)	4 (29%)	9 (56%)	5 (38%)

Table 5.4: Age and Sex demographics of cohort only and allocated trial intervention groups (N= including those randomised at three and at six months)

Within the intervention groups, the manual therapy group were observed to have almost double the proportion of male participants (56%) compared to the cohort and other three intervention groups, which were approximately one third male (29%, 31%, 29% & 38% respectively).

5.3.2 Baseline characteristics for RMDQ and MODI

Table 5.5 shows that the cohort participants had lower RMDQ and MODI, with a mean score of 1.8 and 11.6 respectively, compared to the nested RCT group with a combined score of 8.8 and 28.5.

Objective measure	Cohort only	Nested RCT only
Roland Morris Disability questionnaire (0-24, 0=best[1])	1.8 (2.6)	8.8 (4.2)
Modified Oswestry Disability Index (0-50, 0=best[2])	11.6 (9.7)	28.5 (12.3)

Table 5.5: Mean score of outcome measures at baseline of cohort and nested RCT ([1] RMDQ questionnaire rating scale, [2] MODI questionnaire rating scale)

With minimum entry level into the RCT set a 4 or above on the RMDQ, but no minimum entry level for the cohort group, the cohort presented with a considerably lower score than the minimum eligibility level score for the RCT.

Table 5.6 shows the usual care group had the highest score for the RMDQ with a mean score of 11.4. The usual care group equally had very close to the highest score for the MODI of 29.5. The combined intervention group had the lowest RCT group score for the RMDQ and the MODI at 7.0 and 19.2 respectively.

A between-group difference of more than four points on the RMDQ (7.0 – 11.4) and more than ten points on the MODI (19.2 – 29.6) was observed.

The Acupuncture and manual therapy group had very similar scores on the RMDQ of 8.8 and 8.0 respectively, but a 5.6 difference in the MODI of 29.6 and 24.0 respectively.

Quality of life measure	Cohort study (N=30)	Usual care (N=16)	Acupuncture (N=14)	Manual therapy (N=16)	Combined (N=13)
Roland Morris questionnaire (0-24, 0=best)	1.8 (2.6)	11.4 (5.2)	8.8 (4.3)	8.0 (4.4)	7.0 (2.6)
Modified Oswestry (0-50, 0=best)	11.6 (9.7)	30.3 (14.9)	29.6 (12.2)	24 (13.6)	19.2 (8.0)

Table 5.6: Mean score of outcome measures at baseline of cohort and allocated trial intervention groups *(N= including those randomised at three and at six months)*

The Acupuncture group scored 8.8 on the RMDQ, 2.6 points lower than the usual care group, but held the highest MODI score at 29.6. Only the acupuncture group demonstrated a variation between its RMDQ and MODI scores.

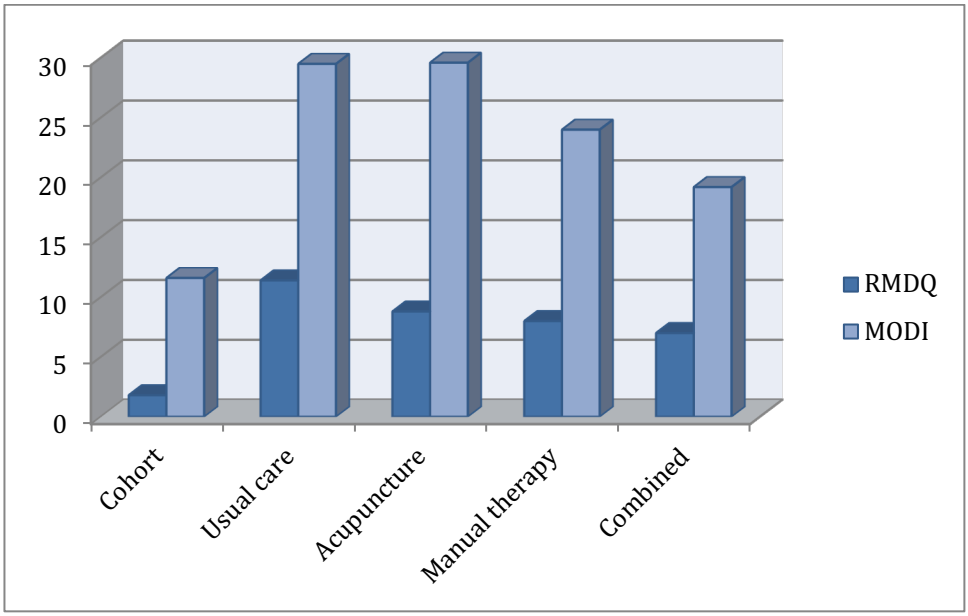

Figure 5.9: Baseline RMDQ and MODI group scores

5.4 Exploratory analysis

A preliminary investigation into the effectiveness of the two interventions (acupuncture and manual therapy) was conducted. Exploratory regression statistical analysis was conducted using two-sided significance at the 5% level on an intention-to-treat basis to consider any noteworthy findings.

For the two participants who were randomised at six months from the cohort study, their six-month score has been classed as their screening score; their three (nine actual) month follow-up data was not available within the time frame of this thesis.

5.4.1 Pre and post-test correlation of Roland Morris Disability Questionnaire

Figure 5.10 shows a scatterplot of pre and post-test values and correlation of the RMDQ scores.

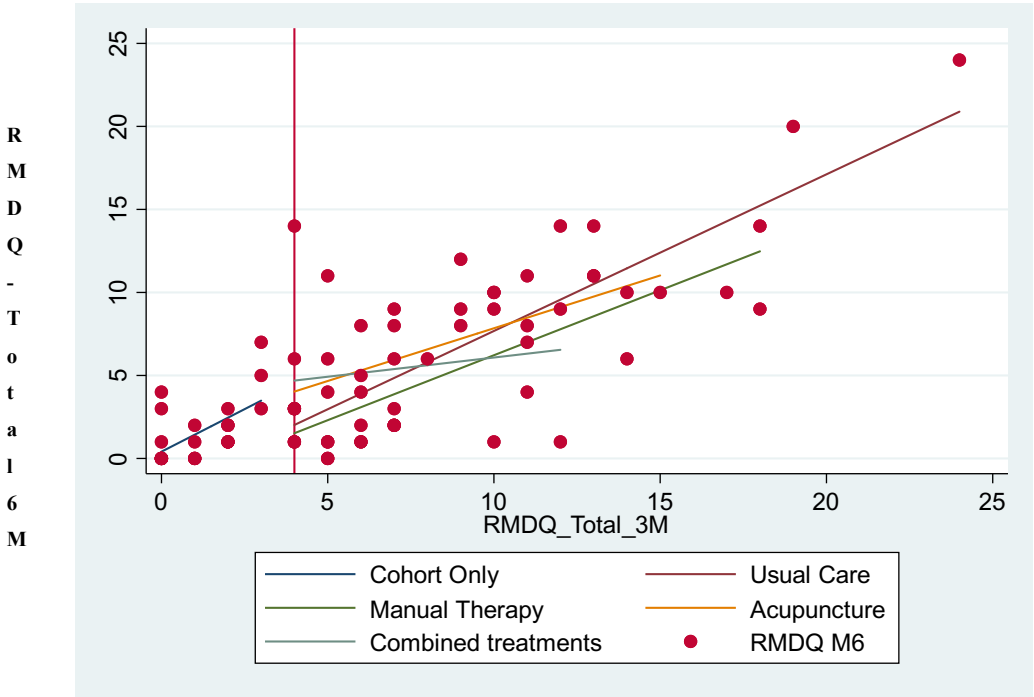

Figure 5.10: Pre and post-test correlation of Roland Morris Disability Questionnaire (RMDQ) (Dascanio, 2014)

Figure 5.10 plots the screening scores at three months (trial entry time point) against the scores three months later at six months (primary analysis). It also shows regression lines for the cohort only group (far left) and for each of the trial intervention arms. The lines represent best fit for the cohort and each of the randomised groups. The vertical line at four represents the trial eligibility entry level on the RMDQ, as participants had to cross this threshold to be eligible for the treatment trial.

A 45-degree line shows where 3month and 6 month scores are the same (no change in symptoms). The lines with a gradient below 45 degrees show improvement (six month scores are less than the three month scores). A higher score of the RMDQ indicates a worse outcome. Lines of positive gradient (>0°) indicate that, in general, the higher a participant scored at month three the higher their score at month six (e.g. a participant who scored 10 at month three tended to score higher at month six than someone who scored 5 at month three). All the regression lines have a positive gradient. The regression line for the cohort has a gradient that appears to be almost 45° indicating that within this group participants' low back pain scores tended to be fairly consistent at three and six months. For the randomised groups, the usual care line is the steepest (i.e. less improvement in symptoms) whilst the combined therapy line is the shallowest (greatest improvement in symptoms). This means that for every unit increase in month three score the increase in predicted month six score was smaller for patients allocated to the combined therapy group than for those allocated to the acupuncture group, the manual therapy group, or the usual care group.

5.4.2 *Exploratory analysis - Roland Morris Disability Questionnaire (RMDQ)*

Quality of life measure	Usual care	Acupuncture	Manual therapy	Combined
Roland Morris questionnaire Baseline	11.4 (5.2) (N=16)	8.8 (4.3) (N=14)	8.0 (4.4) (N=16)	7.0 (2.6) (N=13)
Roland Morris questionnaire Follow-up	9.5 (6.3) (N=14)	6.8 (4.5) (N=13)	4.6 (4.0) (N=15)	5.4 (4.8) (N=11)

Table 5.7: Mean outcome measure scores at baseline and follow-up
(N= at baseline includes participants randomised at three and six months; N= at follow-up excludes those randomised at six months, three selective participants not included in analysis and one incomplete data set)

Observations at baseline and follow-up (table 5.7) show all arms improved in the RMDQ.

In table 5.7 a variation in participant numbers is observed from baseline (59 participants) to the follow up (53 participants). There are three plausible explanations for this. Firstly due to two participants being unwilling to receive acupuncture and thus were randomized only to either usual care or manual therapy and one participant being unwilling to receive manual therapy and thus randomized only to either usual care or acupuncture. The data from these three participants were excluded from any exploratory comparisons between acupuncture and manual therapy (alone or in combination). Secondly two patients were randomised to the trial later in the study at the 6 month time point, their 6-month score has been classified as their screening score; this means however that their follow-up data for these participants is not available. The final participant data lost from this observation occurred due to incomplete section of the participants questionnaire, attempts to contact the participant to retrieve this data were made by telephone however the data was not retrieved.

Table 5.8 shows the results of regression analysis of treatments for low back pain at three months post randomisation time point. The analysis of covariance estimates are adjusted for screening score and highlight the differences between groups and also additional difference attributed over usual care. Negative differences represent a favourable outcome for the relevant intervention over usual care.

Exploratory analysis of the efficacy of the trial interventions showed that the Roland Morris questionnaire scores improved across all the groups.

QoL measure	***Usual care (UC)***	***Acupuncture***	**Additional difference attributed to acupuncture over usual care (95% CI)**	***Manual therapy***	**Additional difference attributed to manual therapy over usual care (95% CI)**	***Combined Acupuncture and manual therapy***	**Additional difference attributed to combined therapy over usual care (95% CI)**
RMDQ	9.3 (6.3) N=14	6.8 (4.5) N=13	**0.3 (-2.9, 3.5) P=0.85**	4.6 (4.0) N=15	**-1.4 (-3.8, 1.0) P=0.24**	5.4 (4.8) N=11	**0.01 (-4.3, 4.3) P=1.00**

Table 5.8: Regression analysis results for RMDQ at 6 months (3/12 post randomisation)

Table 5.8 indicates the manual therapy group produced the largest benefit over usual care at three-months. In the 'additional difference' columns negative differences reported represent a favourable outcome for the relevant intervention over usual care due to the change achieved.

The manual therapy group achieved the greatest negative difference (additional reduction in RMDQ relative to usual care) of -1.4 (95% CI -3.8, 1.0, P=0.24) and thus a favourable outcome above usual care at three months was seen, although statistical significance was not achieved.

Positive numbers reported in the additional difference columns (table 5.8) represent no difference to usual care; a favourable outcome over usual care was not seen. Acupuncture was reported at 0.6 (p=0.78) and combined therapy at 0.01 (p=1.00). Statistical significance was not achieved.

No additional benefit in mean RMDQ score was observed in the acupuncture and combined therapy arm at three months over the usual care group, statistical significance was not achieved due to the low power of the trial to detect such a change, though this was not the objective of the pilot study.

Compared to baseline scores participants in the combined intervention group experienced an increase on average a 1.4-point (95% CI -1.5 to 4.4, p=0.31) in Roland Morris score than the manual therapy group, though statistical significance not achieved.

Compared to baseline scores participants in the combined intervention experienced a 0.9-point (95% CI: -2.9 to 4.7, p=0.63) greater improvement than the acupuncture group after adjusting for screening scores. Statistical difference was not achieved between the combined and acupuncture comparison.

Analysis of an interaction within the factorial design was not conducted for this pilot due to low participant numbers, but a full analysis would be planned for a full-scale trial.

5.4.3 Exploratory analysis - Modified Oswestry Disability Index

Quality of life measure	Usual care	Acupuncture	Manual therapy	Combined
Modified Oswestry Baseline	30.3 (14.9) (N=16)	29.6 (12.2) (N=14)	24 (13.6) (N=16)	19.2 (8.0) (N=13)
Modified Oswestry Follow-up	29.2 (21.0) (N=13)	25.4 (12.0) (N=13)	18.3 (11.1) (N=15)	16.7 (10.9) (N=11)

Table 5.9: Mean outcome measure scores at baseline and follow-up

(N= at baseline includes participants randomised at three and six months; N= at follow-up excludes those randomised at six months, three selective participants not included in analysis and two incomplete data sets)

Observations at baseline and follow-up in table 5.9 show that the MODI improved in all arms. A variation in participant numbers is observed from baseline (59 participants) to the follow up (52 participants). Three explanations are presented in section 5.4.2, however one additional participant data was incomplete for the MODI questionnaire, thus missing data for two participants in total. Attempts to contact the participants to collect this data were made by telephone without success.

Table 5.10 shows the results of regression analysis of treatments for low back pain at three months post randomisation time point. Highlighting the differences between groups and also additional differences attributed over usual care. In the additional difference columns within table 5.10, negative differences reported represent a favourable outcome for the relevant intervention over usual care.

The exploratory analysis of the efficacy of the trial interventions after adjusting for baseline screening score showed (as with the RMDQ), that the modified Oswestry (MODI) scores, the greatest difference was observed in the Manual therapy (additional improvement to usual care of -5.0 points (95% CI -13.3, 3.3, P=0.23) (Table 5.10) though statistical significance was not achieved. The study was not powered to be able to

find statistically significant differences. The difference observed for the manual therapy group appeared consistent across the MODI and the RMDQ.

QoL measure	***Usual care (UC)***	***Acupunc-ture***	**Additional difference attributed to acupuncture over usual care (95% CI)**	***Manual therapy***	**Additional difference attributed to manual therapy over usual care (95% CI)**	***Combined Acupuncture and manual therapy***	**Additional difference attributed to combined therapy over usual care (95% CI)**
MODI	29.2 (21.0) N=13	25.4 (12.0) N=13	**-1.9 (-9.9, 6.1) P=0.63**	18.3 (11.1) N=15	**-5.0 (-13.3, 3.3) P=0.23**	16.7 (10.9) N=11	**2.1 (-6.5, 10.6) P=0.62**

Table 5.10: Regression analysis results for MODI at 6 months (3/12 post randomisation)

Acupuncture achieved a negative difference of -1.9 (p=0.63) and thus a favourable outcome above usual care, though statistical significance was not achieved.

For the combined treatment group no substantial differences were observed after the three month time period, after adjusting for baseline screening score a positive score of 2.1 (-6.5, 10.6. P=0.62) was observed (table 5.10). Positive numbers reported in the additional difference columns within table 5.10 show no favourable outcome over usual care. A difference attributable above usual care was not observed for the combined intervention. Statistical significance was not observed due to the low power of the trial.

The combined arm experienced on average a 2.2 points (95% CI: -4.8, 9.3; P = 0.52) increase in MODI than the manual therapy group and a 0.5point (95% CI: -7.7, 8.7; P = 0.10) increase over the acupuncture arm after adjusting for screening scores, statistical difference was not achieved.

Both differences in the MODI favour the individual intervention rather than the combined group.

Analysis of an interaction within the factorial design was not conducted for this pilot due to low participant numbers, but a full analysis may be planned for a full-scale trial.

5.4.4 A comparison and analysis of the RMDQ and the MODI outcome measure tools

Both outcome measures are frequently used in research and both have been shown to be valid and reproducible, however they each have strengths and limitations in different aspects (Longo et al, 2010).

Both measures were completed fully by 97% of participants returning their questionnaires and all participants appeared happy to complete both the measures, suggesting it is possible for two outcome measures to be used within a trial.

The objective measure tools were compared to each other (see table 5.11).

No significant differences between the two outcome measures data were observed in this pilot study. Both measures were consistent in showing that the manual therapy intervention arm was superior to the other arms; this result was not statistically significant however.

With the RMDQ being scored out of 24 and the MODI scored out of 50 the observed difference between four points and ten points is effectively very similar in observed measurement on the objective measure tools across the study arms, generally indicating the alignment of the outcome measurement scores, showing comparability at baseline. Comparison of the scores using standardised mean differences is shown in section 5.4.5 and would need to be conducted if considered for a full-scale trial.

QoL measure	*Usual care (UC)*	*Acupunc-ture*	**Additional difference attributed to acupuncture over usual care (95% CI)**	*Manual therapy*	**Additional difference attributed to manual therapy over usual care (95% CI)**	*Combined Acupuncture and manual therapy*	**Additional difference attributed to combined therapy over usual care (95% CI)**
RMDQ	9.3 (6.3) *N=14*	6.8 (4.5) *N=13*	**0.3 (-2.9, 3.5) P=0.85**	4.6 (4.0) *N=15*	**-1.4 (-3.8, 1.0) P=0.24**	5.4 (4.8) *N=11*	**0.01 (-4.3, 4.3) P=1.00**
MODI	29.2 (21.0) *N=13*	25.4 (12.0) *N=13*	**-1.9 (-9.9, 6.1) P=0.63**	18.3 (11.1) *N=15*	**-5.0 (-13.3, 3.3) P=0.23**	16.7 (10.9) *N=11*	**2.1 (-6.5, 10.6) P=0.62**

Table 5.11: Regression analysis results, three months post randomisation of interventions and difference attributable over usual care for LBP

5.4.5 Differences between the two objective measures

As this was a pilot study, both of the above measures were used to investigate which would be the more informative measure to use in a future full-scale study of LBP.

A comparison of the two questionnaires performances to assess their reliability to each other as similar measures for LBP showed some similarities and differences.

All groups improved for both measures. Both measures performed comparably for the manual therapy intervention group with an attributed difference above usual care, both providing a negative result within table 5.11. By dividing the negative result (-1.4, -5.0) by the SD the standardised difference can be observed between the two measures. The RMQD thus has a standardised measure of 0.35 and the MODI standardised measure is 0.45, showing minimal difference between the two

measures, with the MODI being slightly more sensitive, however this study was not powered to test for a statistical difference between the measures.

Comparing the results a negative value = a positive effect. Negative values were achieved for the manual therapy group for both measures (RMDQ -1.4, MODI -5.0). However participants in the acupuncture group had a greater improvement on the MODI measure than the RMDQ. A negative value was achieved for the MODI measure (-1.9) but was not achieved for the same participants completing the RMDQ (0.3).

After adjusting for screening scores, consideration of the RMDQ only would suggest the manual therapy group could be effective comparatively, with a small difference in favour of acupuncture and the combined group. Consideration of the MODI only would also indicate the manual therapy group to be superior, with the acupuncture group achieving a difference attributable over usual care and no additional difference observed in the combined therapy group over usual care.

While the results of the pilot are not powered to detect differences it is interesting to observe a different emphasis between the two measures with RMDQ = favouring Manual therapy and the MODI = favouring Manual therapy and acupuncture. Statistical significance was not achieved for these comparisons.

5.4.6 Ancillary data

Descriptive data for the outcome measures SF12, EQ5D, VAS pain scale and the cost analysis questionnaire are presented below. StataSE (64-bit) was used to analyse the results. Mean and standard deviations (or percentage measures) are shown.

For the SF12 (PCS (5.12) and MCS (5.13) analysis), EQ5D (THERM) and the VAS pain scale a loss of one to two participants' data was observed across the trial groups for all the analyses, this derived from incomplete

sections of the participant questionnaires. The recovery of missing data for these outcome measures was not conducted due to the time lapse for analysis of these measures for this PhD, it would be useful in a future trial to plan for the collection of any missing data.

Table 5.12 below shows the PCS scores for the quality of life measure SF12 at baseline (three months) and follow-up (six months).

SF12 PCS Quality of life measure	**Cohort**	**Usual care**	**Acupuncture**	**Manual therapy**	**Combined**
Baseline - three months Pre-randomisation Screening score Mean (SD)	52.8 (4.96) N=27	42.6 (8.68) N=16	43.4 (9.18) N=14	46.5 (12.68) N=16	48.1 (4.53) N=13
Follow up - six months Mean (SD)	51.3 (5.95) N=25	41.9 (10.43) N=15	43.7 (8.73) N=13	48.5 (12.0) N=14	51.9 (6.7) N=12

Table 5.12: SF12 PCS mean score (SD) at baseline (three months) and six months of the cohort and allocated trial intervention groups

Minimal changes in the data are observed from baseline to follow up for the PCS of the SF12. For the combined group an increase of 3.8 points was observed and for the manual therapy group an increase of 2 points was observed. A 0.3 point increase was observed for the acupuncture group and the usual care and cohort groups decreased by 0.7 and 1.5 respectively. A small improvement was observed in the combined and manual therapy groups and a small deterioration was observed in the usual care and cohort groups; however the sample size was small thus limiting any interpretation from the results.

Table 5.13 shows the MCS scores for the quality of life measure SF12 at baseline (three months) and follow-up (six months).

SF12 (MCS) Quality of life measure	Cohort	Usual care	Acupuncture	Manual therapy	Combined
Baseline - three months Pre-randomisation Screening score Mean (SD)	50.7 (8.57) N=27	43.7 (11.13) N=16	43.9 (9.89) N=14	47.3 (12.49) N=16	52.6 (7.16) N=13
Follow up - six months Mean (SD)	52.0 (7.99) N=25	45.2 (10.1) N=15	50.8 (9.06) N=13	51.1 (8.59) N=14	49.9 (7.94) N=12

Table 5.13: SF12 MCS mean score (SD) at baseline (three months) and six months of the cohort and allocated trial intervention groups

The greatest increase was observed in the acupuncture group, with a 6.9 point score increase. While the other groups observed a more modest increase of 3.8 points for the manual therapy group, a 1.5 points for the usual care group and a 1.3 points increase for the cohort group. The combined group observed a decrease in score of -2.7 points suggesting a deterioration in MCS score, however the sample size was small thus limiting any interpretation from the results.

Table 5.14 shows the scores for the quality of life measure EQ5D at baseline (three months) and follow-up (six months).

For the EQ5D, no difference was observed for the cohort and usual care groups at baseline and follow up. A 1.5 point increase was observed in the acupuncture group, a 1 point increase in the combined group and a 0.5 point increase in the manual therapy group, with an increase in the score showing a mild improvement from baseline to follow up, however the sample size was small thus limiting any interpretation of the results.

The THERM data showed a reduction in score for the cohort and usual care group of 4.6 and 2.4 points respectively. However an increased score was observed for the acupuncture (10.2 points) manual therapy (7.4 points)

and the combined group (1.5 points) suggesting a deterioration of the THERM score for participants, however the sample size was small thus limiting any interpretation of the results.

EQ5D Quality of life measure	**Cohort study**	**Usual care**	**Acupuncture**	**Manual therapy**	**Combined**
Baseline - three months Pre-randomisation Screening score Mean (SD)	0.8 (0.18) N=27	0.6 (0.28) N=16	0.6 (0.21) N=14	0.8 (0.14) N=16	0.7 (0.10) N=13
THERM Scale (0-100, 0 =best) Mean (SD)	84.3 (8.56) N=26	69.5 (13.92) N=15	60.5 (19.35) N=14	68.3 (18.64) N=16	77.7 (10.14) N=13
Follow up - six months Mean (SD)	0.8 (0.12) N=26	0.6 (0.32) N=15	0.75 (0.13) N=13	0.85 (0.12) N=15	0.8 (0.23) N=12
THERM (0-100, 0 =best) Mean (SD)	79.7 (16.5) N=26	67.1 (18.27) N=14	70.7 (16.92) N=13	75.7 (20.93) N=15	79.2 (19.4) N=12

Table 5.14: EQ5D mean score (SD) at baseline (three months) and six months of the cohort and allocated trial intervention groups

A considerable improvement of pain rating on the VAS scale was observed for the manual therapy group with a reduction from 33.7 to 14.6 points across the time points, a reduction of 19.1 on the scale. A reduction of 8.6 points for the acupuncture group, 5.4 points for the combined group and 1.4 points for the cohort were observed. An increase in pain score was observed in the usual care group of 4.7 points.

Table 5.15 below shows the scores for the VAS (Visual analogue scale) for Pain at baseline (three months) and follow-up (six months).

VAS Pain Scale (0-100 100 = worst)	**Cohort**	**Usual care**	**Acupuncture**	**Manual therapy**	**Combined**
Baseline - three months **Pre-randomisation Screening score** **Mean (SD)**	27.4 (20.39) N=26	48.3 (21.4) N=15	38.1 (18.54) N=13	33.7 (22.0) N=16	29 (15.34) N=13
Follow up - six months **Mean (SD)**	26 (21.91) N=25	53 (16.4) N=13	29.5 (20.54) N=13	14.6 (11.65) N=15	23.6 (22.5) N=11

Table 5.15: VAS of pain, mean score (SD) at baseline (three months) and six months of the cohort and allocated trial intervention groups

Resource use data: Table 5.16 and table 5.17 below show the results of the EcoCost questionnaire at baseline (three months) and follow-up (six months), mean score (number of visits) and SD's (or percentages) are presented.

Eco Cost Questionnaire at 3 months	Cohort study	Usual care	Acupuncture	Manual therapy	Combined
1) Care from GP in the last three months					
Own or other GP **Mean (SD)**	0.8 (1.31) N=27	1.5 (2.53) N=15	1.4 (1.98) N=14	0.9 (1.11) N=16	0.9 (1.11) N=13
Practice nurse **Mean (SD)**	0.2 (0.5) N=27	0.6 (0.86) N=13	0.7 (1.18) N=13	0.4 (0.63) N=15	0.2 (0.40) N=11
Physiotherapist **Mean (SD)**	0.1 (0.42) N=27	0 (0) N=13	0.4 (0.99) N=12	0.6 (1.59) N=15	0.4 (0.92) N=11
Other **Mean (SD)**	0 (0) N=25	0.6 (1.28) N=14	0 (0) N=10	0.2 (0.59) N=13	0.4 (0.92) N=11
Other detail	-	1x Counsellor 1x Nurse specialist 1x Sports therapist	-	1x Health care assistant 1xMES clinic	1x Chiropractor 1x MRI scan 1x physiotherapist
Other **Mean (SD)**	0 (0) N=25	0.25 (0.86) N=12	0 (0) N=10	0 (0) N=13	0 (0) N=10
Other detail	-	1x Podiatrist	-	-	
2) Care from NHS Hospital - In last three months as an emergency					
Yes (%)	1 (3.7)	1 (6.25)	0	0	0
(No of admissions)		1			
No (%)	26 (96.3)	15 (93.75)	14 (100)	16 (100)	13 (100)
3) Care from NHS hospitals - In last three months non-emergency					
Yes (%)	2 (7.4)	1 (6.25)	0	0	0
(No of admissions)	2	1	-	-	-
No (%)	25 (92.6)	15 (93.75)	14 (100)	16 (100)	13 (100)
4) Care from NHS hospital by a doctor in outpatient clinic					

Doctor **Mean (SD)**	0.11 (0.42) N=27	0.06 (0.25) N=16	0 (0) N=14	0.36 (0.84) N=14	0.15 (0.37) N=13
5) Care from NHS hospitals –In last three months from any other health professional					
Physiotherapist **Mean (SD)**	0.48 (1.19) N=25	0 (0) N=15	0.5 (1.4) N=14	0.94 (2.17) N=16	0.23 (0.6) N=13
Other **Mean (SD)**	0.19 (0.69) N=26	0.53 (1.24) N=15	0.08 (0.28) N=12	0.13 0.35 N=15	0 (0) N=11
Other detail	1x Dietician	1x Counsellor 1x Nurse specialist 1x Podiatrist	1x MRI pelvic scan	1x Podiatrist 1x Urologist	-
Other **Mean (SD)**	0 (0) N=25	0 (0) N=14	0 (0) N=12	0.13 (0.51) N=15	0 (0) N=11
Other detail	-	-	-	1x Consultant Urologist	-
6) Private treatments –In the last three months have you been admitted to a private hospital					
Yes (%)	1 (3.7)	1 (6.25)	0	0	0
No. of admissions	1	1	0	0	0
No (%)	26 (96.3)	15 (93.75)	14 (100)	16 (100)	13 (100)
7) In the last three months how often have you consulted a private healthcare professional					
Doctor **Mean (SD)**	0.38 (0.19) N=26	0.63 (1.63) N=16	0 (0) N=14	0 (0) N=15	0 (0) N=12
Physiotherapist/ Chiropractor/Osteopath **Mean (SD)**	1 (3.16) N=27	0.33 (1.29) N=15	0.57 (2.14) N=14	0.25 (0.77) N=16	0.69 (1.18) N=13
Other **Mean (SD)**	0.25 (0.73) N=24	0.2 (0.77) N=15	0.41 (1.44) N=12	0 (0) N=14	0.08 (0.28) N=13
Other detail	1x Occupational health Physiotherapist 2x Dentist 1x Surgeon	1x Sports therapist	1x Chinese acupressure	-	1x MRI scan knees

Table 5.16: Eco cost questionnaire mean score (SD) and percentage results at baseline (three months) of the cohort and allocated trial intervention groups

Table 5.17 below show the results of the EcoCost questionnaire at follow-up (six months), mean score (number of visits) and SD's (or percentages) are presented.

Eco Cost Questionnaire at 6 months	Cohort	Usual care	Acupuncture	Manual therapy	Combined
1) Care from GP in the last three months					
Own or other GP Mean (SD)	0.57 (0.95) N=26	1.07 (1.70) N=13	0.84 (1.21) N=13	1.35 (1.69) N=14	0.57 (1.62) N=12
Practice nurse Mean (SD)	0.73 (1.93) N=23	0.33 (0.49) N=12	0.27 (0.65) N=11	0.1 (0.62) N=10	0 (0) N=11
Physiotherapist Mean (SD)	0.66 (3.1) N=21	0 (0) N=11	1.36 (3.1) N=11	0.36 (1.20) N=11	3.6 (5.39) N=11
Other Mean (SD)	0.1 (0.44) N=20	0.1 (0.31) N=10	0 (0) N=11	0.25 (0.46) N=8	0.13 (0.35) N=8
Other detail	1x Osteopath	1x Consultant	-	1x asthma check 1x Health care assistant	1x Eye consultant
Other Mean (SD)	0 (0) N=17	0 (0) N=10	0 (0) N=9	0.71 (1.89) N=7	0 (0) N=7
Other detail	-	-	-	No detail	-
2) Care from NHS Hospital - In last three months as an emergency					
Yes (%)	1 (3.85)	1 (7.14)	0	0	0
(No of admissions)	1	1			
No (%)	25 (96.15)	13 (92.86)	13 (100)	15 (100)	12 (100)
3) Care from NHS hospitals - In last three months non-emergency					
Yes (%)	1 (3.85)	1 (7.14)	0	0	0
(No of admissions)	1	1	-	-	-
No (%)	25 (96.15)	13 (92.86)	13 (100)	15 (100)	11 (100)
4) Care from NHS hospital by a doctor in outpatient clinic					
Doctor Mean (SD)	0.23 (0.82) N=26	0.08 (0.28) N=12	0.15 (0.38) N=13	0.6 (1.12) N=15	0.25 (0.62) N=12

5) Care from NHS hospitals –In last three months from any other health professional					
Physiotherapist Mean (SD)	0.53 (2.74) N=26	0.28 (1.06) N=14	0.23 (0.83) N=13	1.54 (2.57) N=13	0.24 (1.44) N=12
Other Mean (SD)	0.87 (0.29) N=23	0.38 (1.12) N=13	0.09 (0.3) N=11	1.1 (1.49) N=10	0 (0) N=9
Other detail	1x Gynaecologist 1x Hypnotherapist	1x Consultant Gynae 1x Counsellor	1x Consultant Rheumatologist	1x Clinical psychologist 1x Gastroenterology 1x Medical elective 1x Moxiofacial 1x X-ray nurse	-
Other Mean (SD)	0 (0) N=22	0 (0) N=11	0.09 (0.30) N=11	0.25 (0.46) N=8	0 (0) N=9
Other detail	-	-	1x Gastroscopy	1x MRI nurse 1x Pain clinic	-
6) Private treatments –In the last three months have you been admitted to a private hospital					
Yes (%)	1 (3.85)	1 (6.67)	0	1 (6.67)	0
No. of admissions	1	1	0	1	0
No (%)	25 (96.15)	14 (93.33)	13 (100)	14 (93.33)	12 (100)
7) In the last three months how often have you consulted a private healthcare professional					
Doctor Mean (SD)	0.12 (0.33) N=25	0.07 (0.26) N=14	0 (0) N=12	0.64 (2.13) N=14	0.38 (1.06) N=8
Physiotherapist/ Chiropractor/Osteopath Mean (SD)	0.29 (0.8) N=24	0.28 (1.07) N=14	1.64 (3.2) N=11	0.71 (2.67) N=14	3.72 (4.17) N=11
Other Mean (SD)	0.25 (0.74) N=24	0 (0) N=14	1.27 (3.13) N=11	0 (0) N=11	0.11 (0.33) N=9
Other detail	1x Acupuncture 2x Chiropodist 1x Sports massage	-	2x Acupuncture	-	1x Eye consultant 1x no detail

Table 5.17: Eco cost questionnaire mean score (number of visits), (SD) and percentage results at six month follow up of the cohort and allocated trial intervention groups

This study was intended to pilot the collection of data from the resource use questionnaires, thus limited information can be determined from the differences between time points due to the small sample size.

Resource use reported in tables 5.16 and 5.17 was observed to be low, across all groups and across both time points. No notable values were observed in the variables data across the tables.

Under the care from GP section, resource use of accessing a GP use increased marginally for the manual therapy group (from 0.9 (1.11) three months, to 1.35 (1.69) six months), but decreased for all other groups from baseline to follow up.

Resource use of accessing a physiotherapist went up across the acupuncture (0.4 to 1.36) and combined (0.4 to 3.6) intervention groups at six months, this occurred due to some individuals mistakenly including their trial intervention sessions on the resource use section of the questionnaire. Individuals indicated they had received 10 physiotherapy sessions, these individuals were contacted by telephone, where possible, to confirm if they had identified the trial interventions on the questionnaire and these were corrected where confirmation was possible. If follow up by telephone was unsuccessful, further follow up was not conducted and data was not corrected. A future study would plan to avoid this inaccuracy in reporting resource use.

Additional resource use across the groups varied but at three months, nine additional contacts in the "other section" where made and included the usual care group accessing a counsellor, nurse, sports therapist and a podiatrist. The manual therapy group accessing health care assistants, the MES clinic, and the combined therapy group accessing chiropractors, MRI scans, physiotherapists. At six months, six additional contacts were made including the cohort accessing an osteopath, the usual care group accessing a consultant, the manual therapy group accessing an asthma check and a health care consultation and the combined group accessing an eye

consultant. The Acupuncture group did not access any additional care in the recorded "other" sections at three or six months.

Hospital use both emergency and non-emergency across the groups was minimal with only one or two participants from the cohort and usual care group only accessing services, at both baseline and follow up.

Variation in completion rates were seen throughout the questionnaire in the resource use section; a loss of one to four participants data was observed across the groups for the analysis at six months, this derived from incomplete sections of the participant questionnaires i.e. all questionnaires were returned but not every question was complete. This figure rose to the loss of up to ten participants data for the sections titled 'other', participants completed part 1,2,3 of the question but left the section 'other' blank. Retrieval of missing data was not conducted.

5.4.7 Effectiveness of the design for evaluating a population with LBP

The design methodology of this study worked effectively in the assessment of interventions for LBP. The details and a discussion of the design will be presented in chapter six.

The factorial design was effective for assessing the use of acupuncture and manual therapy in comparison and in combination while maintaining a control arm. A full-scale trial of this pilot would incorporate the two interventions concurrently and is likely to be a more cost effective trial in its approach, than conducting two separate trials.

5.5 Proposed full-scale study sample size calculation

5.5.1 Calculations to inform sample size calculation

5.5.1.1 Calculation of Standard Deviation (SD):

The SD for the sample size to inform a full-scale study was calculated for the RMDQ score at six months (three months post randomisation) using StataSE (64-bit).

Calculation from this pilot: Standard deviation = 5.0

5.5.1.2 Important difference on RMDQ score

A difference of 1.5 points on the Roland-Morris disability questionnaire would be anticipated for a future study of LBP, this decision was made based on consensus from published literature.

Ellard et al. (2017) observed mean between group differences of 0.8 and 1.87 points in the RMDQ at three months in their meta-analysis of large physiotherapy trials of LBP using the RMDQ. UKBEAM (2004) observed a 1.6-point improvement at three months and 1.0 point at 12 months. While they note it is a small to moderate clinical benefit, it was suggested that this benefit in outcome might lead to a large economic effect due to the high cost of LBP to society (UKBEAM, 2004). A similar finding was found for the pilot study conducted for this thesis for the manual therapy group, the additional reduction in RMDQ relative to usual care was 1.4 points (95% CI: -1.0, 3.8. P = 0.24) at three month follow up, though as it was a small pilot the results were not statistically significance (Dascanio et al. 2019).

Chou et al (2007) in their review of treatments for LBP including acupuncture, found a difference of 2 - 4 on the Roland-Morris disability questionnaire to be significant in the reviewed articles. In a study of yoga for LBP, a difference of 1.48 points lower (95% CI, 0.33 to 2.62 points)

was observed at six months and 1.57 points lower (95% CI, 0.42 to 2.71 points) was observed at 12 months on the RMDQ score (Cox et al. 2010).

The selection of a difference of 1.5 points also equates for the calculation of the planned effect size of 0.3 (section 5.5.1.3). With a planned mean difference of 1.5, divided by the six-month standard deviation for this pilot of 5.0, equalling 0.3 for the effect size:

$$1.5 / 5.0 \text{ (SD)} = 0.3$$

A mean difference of 1.5 points on the Roland-Morris disability questionnaire was therefore used as the mean difference worth being able to detect as statistically significant for the sample size calculation for a potential full-scale study of this pilot.

5.5.1.3 Estimation of effect size:

The effect size is often used for reporting and interpreting effectiveness (Coe, 2002). As this was a pilot study it was not the intention to estimate an effect size from the data collected; with a small sample size the precision of any estimates of effect were of course low. However the calculations have been conducted for exploratory reasons and presented below for interest. The six-month data set has been used, selecting the mean RMDQ score from the acupuncture (for the purpose of this calculation) and usual care groups (table 5.8).

$$\text{Effect Size} = \frac{\text{(Mean of experimental group) – (Mean of control group)}}{\text{Standard Deviation}}$$

$$\text{Effect size} = \frac{\text{(6.8 Acupuncture) – (9.3 Usual care)}}{\text{5.0 (SD)}} = -0.5$$

An effect size of 0.5 was estimated, this calculation was based on the six-month data set for this study (table 5.8).

An effect size of 0.3 is recognised as a medium effect (Cohen, 1992), and I decided to use this to calculate the sample size for a future trial of this study, as it is more conservative. The rationale for selecting a medium effect size of 0.3 is based on the recognised use of a medium effect size for RCTs (Cohen, 1992) and published literature reporting their effect sizes.

Rothwell, Julious and Cooper (2018) reported the median standardised target effect size across the studies was 0.30 (interquartile range: 0.20 – 0.38) from 107 RCTs included in their study. UKBEAM (2004) reported an effect size of 0.39 at three months and 0.25 at 12 months for the manipulation group on the RMQ. MacPherson et al (2014) reported effect sizes of 0.17 (95% CI: 0.11, 0.23) and 0.43 (95% CI: 0.01, 0.85) for the effect of acupuncture versus sham penetrating and non-penetrating needles comparatively across 29 high quality RCTs. Linde K, Niemann K and Meissner K (2010) reported effect sizes of 0.43 (95% CI: 0.01, 0.85) and -0.37 (95% CI: -0.70, 0.04) for the effect of acupuncture versus sham penetrating and non-penetrating needles comparatively across 61 RCTs.

After consideration of the reported effect sizes in the literature, the effect size calculated from this pilot (above) and the planned mean difference of 1.5 on the RMDQ (section 5.5.1.2) an effect size of 0.3 was considered appropriate to select, to calculate the sample size for a future full-scale RCT.

5.5.1.4 Power

A 90% power was selected for calculating the study sample size required to have a high probability of detecting the difference as statistically significant if a difference exists between the interventions.

Conventionally it is recommended to have power of 80% to 90% for an RCT (Hickey et al. 2018), however 90% power requires an increase in sample size of around 30% (Wittes, 2002).

5.5.1.5 Summary of calculations to inform sample size calculation

In summary a mean difference of 1.5 points on the Roland-Morris disability questionnaire would be appropriate for future calculations, with an effect size of 0.3, assuming the standard deviation of 5.

5.5.2 Sample size calculation

A sample size calculation was performed to inform a full-scale trial; using PS Power and sample size calculation version 3.1.2. A future full-scale study would be planned to follow a continuous response variable of independent control and experimental participants.

The pilot was planned to observe a response from each trial group of normal distribution, with a standard deviation of 5, which was observed from the study. If the true difference in the experimental and control means was 1.5, 234 experimental participants and 234 control participants would be needed to be able to reject the null hypothesis; that the population means of the experimental and control groups are equal with probability (power) 0.9.

The type I error probability associated with this test of this null hypothesis is 0.05. Translating this power calculation into a four group factorial design, would relate to 117 participants in each of the four groups of the RCT. As the comparison of active treatment is with usual care, a comparative treatment group and a combined treatment group.

468 participants would therefore be required to detect an effect size of 0.3 with statistical significance at the 5% level for a full-scale randomised controlled trial of this pilot.

While the pilot study observed a post randomisation attrition rate of zero, it is good practice in research to allow for 10% loss due to attrition, the sample size would be inflated by an anticipated attrition rate of 10%. 468

/ 0.9 = 520 participants, this would allow 130 in each of the four trial arms. 520 participants would therefore be required for recruitment and random allocation to one of the four trial groups within a future full scale RCT.

If the cohort recruitment design with nested RCT were to be adopted for the full scale study, based on the pilot study conducted, these numbers would need to inflated by 30%, as 30% of the participants who consented to the study failed to return their baseline questionnaire or subsequently left the study, therefore: 520 / 0.7 = 743. This would allow for recruitment to the cohort and for the loss of participants during the first three month time period, prior to randomization to the RCT, as seen for this pilot and described in section 5.2.4 and 5.2.5.

30% of the cohort participants in the pilot were also ineligible to enter the nested RCT at the three-month time point due to a low score on the RMDQ, as they did not score the entry threshold of 4. If the design remained the same and a minimum RMDQ score was not applied to cohort entry, then a further 30% (743 / 0.7 = 1061) inflation of participants would be required to allow for participants in the cohort being ineligible for the RCT.

Therefore, based on the information derived from this pilot and utilizing the same study design, a target of 1064 participants would be required for recruitment to a future cohort study in order to achieve a target of at least 520 of those participants entering, and 468 completing the nested RCT.

- *468 / 0.9 = 520*
- *520 / 0.7 = 743*
- *743 / 0.7 = 1064 (1061 round up to full group numbers)*

Therefore a sample size of 1064 participants would be required for a full-scale study of this pilot.

5.5.2.1 Interaction and additive effects

A future full-scale trial with a factorial design is unlikely to be planned to test for interaction and / or additive effects of the assessed treatment interventions, due to the considerable increase in sample size required. These numbers are shown below for illustrative purposes.

> *Interaction: (also known as effect modification) occurs when the effect of one variable (intervention) on an outcome is modified according to the level of a second variable (intervention) (Altman and Bland, 2003).*
>
> *Additive effects: occur when two factors or variables (interventions) have an additive effect on each other, adding to the others outcome of effect, without interaction. (Small, Volpp and Rosenbaum, 2011)*

To test for interaction and additive effects then the RCT sample size would need to be inflated by a further multiple of 4, to power the study to effectively.

- *520 x 4 = 2080*

If nested within a cohort design, based on the findings of this pilot, further inflation of 30% to allow for a cohort design and to account for loss pre randomisation.

- *2080 / 0.7 = 2971*

Additional loss when recruitment from the cohort to the RCT for randomisation, thus a further inflation of 30% would be required to account for ineligibility within the cohort prior to entry into the RCT

- *2971 / 0.7 = 4245 (full-scale sample size calculation)*

Therefore 4245 participants would be required for a full-scale study of this pilot for the additional power to test for interaction and additive effects of the interventions.

5.6 Projected full scale study calculation

Based on the pilot study results to achieve a cohort participant rate of 1064 (section 5.5.2) for a future full-scale study, a more extensive trial infrastructure would be required:

GP recruitment:

A future trial would require inviting 20 – 25 GP practices to allow for the successful recruitment of at least 15 of those practices. The number of registered patients would need to be in excess of 262,703 across 9 practices, if a similar identified eligible patient rate of 2.7% and a participant response rate of 15% was projected, to achieve the number of participants required.

From the information gained from the pilot, a mail out to at least 7093 patients would be anticipated to achieve the required 1064 participant numbers, over a six to nine month time period.

Physiotherapy recruitment:

On the basis of recruiting physiotherapy practices similar to the one in the pilot study, if each physiotherapy practice had three physiotherapists, (at least two of whom were qualified to practice acupuncture) and each practice being able to receive and treat 84 participants: nine physiotherapy practices would be required for a future trial. However as was the case with the physiotherapy practice in the pilot it is anticipated that the physiotherapy practices would be able and willing to accommodate additional participants, the pilot physiotherapy practice reported being available to accept at least 135 participants over the course of this trial and they were a relatively small private practice. On the basis of the

information provided from this pilot, five – six physiotherapy clinics would need to be recruited for a future trial.

Therefore 15 physiotherapy practices would need to be contacted and invited to achieve the successful recruitment of five to six. I anticipate that at least two of the recruited practices would be NHS outpatient clinics to provide a balance of NHS and private practice input into the study.

Participant recruitment:

Alongside GP database recruitment, local advertising of the trial should be considered including notices at private clinics, local groups and sports centres to access those individuals who do not report their LBP to their GP. This would provide further information regarding the population suffering with LBP and provide guidance on the extent of under reporting that occurs within society (Papageorgiou, 1991).

Additionally to achieve a participant sample more ethnically and socially diverse and in line with the UKs general population, a multi-regional trial would need to be considered, using at least two regions of the UK (e.g. York and Leeds), this would allow for a potentially more inclusive participant population.

Recruitment from cohort study to nested RCT:

A 65% recruitment rate from the cohort to the treatment trial is a useful figure for informing future studies. It is also a significantly high rate of recruitment comparative to the observed recruitment of two previous studies of between 12 – 15%, who also used GP database recruitment methods but recruited directly into their trial (Cox et al. 2010, Reynolds et al. 2008). An even higher rate of trial recruitment may also have been achieved if the 30% of the cohort who were ineligible for the study had scored higher on the RMDQ, the consideration of having a minimum entry level of the RMDQ to the cohort for a future study may aid a greater cohort to nested trial recruitment rate.

Box 5.1: Summary of key study findings

Key findings

- *The design was effective for recruiting particiapants from GP to the cohort study*
- *The design followed a novel methodology and the cohort was useful for recruiting participants from to the nested RCT*
- *Consent and intervention acceptance rates were 97%*
- *This observational cohort study with a nested factorial RCT demonstrated zero attrition post randomisation*
- *This design with factorial RCT was appropriate for evaluating acupuncture and manual therapy for LBP*
- *The outcome measure tools were appropriate for LBP*

What this adds to what is known:

- *The design has not been previously used in the study of LBP and is an appropriate design for this population*
- *Acupuncture and manual therapy can be combined by physiotherapists safely; it is achievable and potentially effective in combination*
- *Participants willingly accept acupuncture and manual therapy as interventions for LBP*

What is the implication, what should change now:

- *When evaluating novel interventions in chronic musculoskeletal problems, trials should consider using a cohort design with a nested factorial randomised controlled trial*
- *Cost effective delivery of combining interventions should be considered to reduce cost to the NHS*

(Dascanio, 2015d)

5.7 Summary

The aim of this study was to conduct a pilot study to investigating the use of a novel trial design; an observational cohort study with a nested factorial randomised controlled trial, for a population suffering from LBP. The methodology of a cohort study with a nested RCT for participant recruitment would be focused upon.

The pilot study recruitment and attrition rates and the acceptability of acupuncture and manual therapy as a treatment for people with LBP were the key objectives. The results of the study have been presented within this chapter and the design was found to achieve the objectives of the study.

With the sample size and the study not planned to establish effectiveness differences, the differences noted in the study are small to moderate, however on a larger scale we would expect to see further differences across the groups and analysis.

Discussion of the above results analysis and the pilot RCT are presented in chapter six.

6 Discussion of the Pilot Study

6.1 Introduction

In this chapter I discuss the pilot study (chapter four) and the results presented in chapter five. I consider the aims and objectives of the study and how these were achieved. I review the results presented and discuss the strengths and weaknesses of the pilot study design.

The aim was to conduct a pilot to investigate the use of an observational cohort study with a nested factorial RCT methodology, for evaluating acupuncture and manual therapy for LBP. Recruitment, retention and attrition rates and the acceptability of acupuncture and manual therapy as a treatment were the key objectives for the study and these will be discussed further as part of this analysis. Conclusions will be formed, discussing any implications for a future full-scale trial.

The reporting of data for this study constitutes the data from commencement of the study to the six-month data time point and will inform this discussion. Data beyond the six-month time point is outside the remit of this PhD and does not form part of this thesis.

6.2 Recruitment

6.2.1 General Practice (GP) recruitment

Of the five GP practices invited two were recruited over a four-month period. The recruitment of the GP practices was conducted via email, mail, telephone and face-to-face meetings and the time taken to recruit the two GP practices took nearly as long as recruiting the participants to the study. The delays in GP recruitment occurred in part due to unavailability of required personnel, changes in personnel during the process and administrational requirements of the NHS organisations. The paperwork

involved in the GP recruitment, recording of patients contacted and those recruited was time consuming and working towards an automated recording system for a larger trial would be preferential.

As discussed in section 5.6, inviting between 20 – 25 GP practices and recruiting a minimum of 15 GP practices would be required for a full-scale study. The planning for the logistics of this and setting a time frame for recruitment would need to be factored into the planning of a future study.

6.2.2 Physiotherapy practice recruitment

During the recruitment process one physiotherapy practice was recruited to this pilot trial. At the time of recruitment, the single physiotherapy practice employed eight physiotherapy staff at the practice, five were acupuncture trained, providing a variety of physiotherapy clinicians to accommodate all the potential trial participants. Though only one practice was recruited the staffing levels were considered sufficient for the pilot trial.

Subsequently however by the time of the commencement of the trial interventions, the practice faced some unexpected external contract changes and staff redundancies, leaving three staff members remaining, two that were able to practice acupuncture. Though this was a relatively small number of therapists and a greater number would have been preferable, having three therapists did reduce the risk of the results being dependent upon a single therapist's expertise and any practitioner effect bias would have been eliminated. Therefore, it was determined at this stage, that this was still adequate to accommodate the trial participants for the purpose of this pilot study.

For a full-scale trial a greater ratio of physiotherapy staff would be intended, desirable and best practice to deliver the interventions. It is important to have more than one physiotherapist to deliver the trial interventions to eliminate any risk of the results being dependent upon a practitioner or therapist effect and the introduction of bias.

The recruitment of more physiotherapy practices to limit the effects of any staff changes at the private practices impacting on the trial would be planned. It would also be appropriate to use both NHS and private practice physiotherapy services for a future full-scale study to more closely reflect societal norms. Confidence that one could recruit sufficient physiotherapy practices can be gained from various studies. UKBEAM (2004) scaled up their feasibility study successfully and used both the NHS and private practitioners concurrently for their study of manipulation and exercise for LBP. Vickers et al. (2002) reported acupuncture is widely used in both private practice and the NHS. A mapping survey of acupuncture services in the UK estimated four million acupuncture sessions are conducted annually, 330 practitioners responded (29% Doctors, 29% physiotherapists, 15% nurses, 27% acupuncturists) of these 68% provided services in the private setting and 42% practiced in the NHS (Hopton et al. 2012). As discussed in section 5.6 having a minimum of five physiotherapy practices (with at least two NHS sites) with multiple staff in each would be recommended for a full-scale study.

Additional ethics approval would be needed for any included NHS sites. Further consideration of participant logistics (access, parking at NHS sites) would need to be considered. NHS physiotherapists' may have a requirement for additional training, resource provision, available time for trial participants and diary availability, (i.e. diary availability for rebooking participants at follow-up within one week) due to their often overstretched caseloads within NHS sites. The pilot study only demonstrated the successful recruitment of one physiotherapy practice and consideration would need to be given to recruit additional practices. The national Primary Research Networks would be a useful resource for supporting the recruitment of NHS physiotherapy sites, and advertising and direct contact with geographically suitable private practices would be planned. Further discussion of scaling up the pilot study, the recruitment of physiotherapists and the inclusion of NHS sites are provided in section 8.6.

6.2.3 Participant recruitment

Recruitment of participants occurred over a six-month period from the two GP surgeries. This was a reasonable time frame for sufficient recruitment of participants to the pilot cohort study. A greater time frame would need to be considered to accommodate a larger scale study and there would need to be consideration of the benefit of additional GP practices if it increased the recruitment period substantially. It may be that with a wider geographical site and additional physiotherapy practices, more than one GP practice and recruitment could be conducted simultaneously. Only having one physiotherapy practice for this study limited the number of participants that could be admitted at one time.

6.2.4 GP database recruitment

The difference between the total of identified potential participants across the two recruited GP surgeries was notable. GP practice one; with a greater population of 17564, yielded 275 (1.6%) potentially eligible LBP participants in their search, whereas GP practice two with a significantly smaller population of 13538 identified 570 (4%) potentially eligible LBP participants. There was a distinct difference between the percentages of recruitment of participants with LBP between the two GP practices at with the smallest practice generating a significantly greater number of potential participants. 2.7% of patients on the total GP databases were recorded to have LBP.

Further consideration as to whether this was due the GP database software systems and their algorithms to identify potential participants or whether the characteristics of the population of the practice differed are unclear from this small sample. An exploratory investigation to gain knowledge and understanding of the identification process of all the GP database systems in the area would be useful to inform future researchers, to enable the maximisation of results from this recruitment model. It would enable the identification of the most effective systems to utilise, the provision to

guide expectations, and also indicate the number of GP surgeries required for future recruitment in studies.

The divergence in numbers identified across the two practices may also have been due to the differing demographics of the population in each GP region reporting LBP. There is little to suggest any differing demographics of the York regions used or any vast difference in the cultural norms. However due to the demographics of the York region there was generally limited cultural diversity amongst the participants within this pilot study. Extending the region would need to be a consideration for a full-scale trial, to ensure the trial population is representative and generalisable to the whole of the UK.

Additional examination of the area demographic facilities may have been of value prior to the study. Demographic factors may have influenced the numbers of participants reporting LBP to their GPs. In some area's patients may seek private treatment directly or have access to a direct access to physiotherapy via a self-referral pathway without ever consulting their GP regarding their LBP. The presence of an effective private therapist in the region or direct access to physiotherapy may influence reporting processes and conversely these services may not be available in some areas.

As discussed in chapter one, LBP is a vastly under reported condition with only approximately 20% of sufferers reportedly consulting their GP (Papageorgiou, 1991). More recent epidemiological data on the reporting of LBP to a GP is presently unavailable, a UK National Pain Audit was published in 2012 (Price et al. 2012) but no further information on GP reporting rates was provided. It may be that LBP is significantly under reported at some GP practices compared to others. Recruitment of participants via other methods, through advertising, employers and from other local therapists would need to be a consideration. Further demographic planning and discussion would aid and inform a future study.

6.2.5 Recruitment rates of participants to the cohort form GP practices

Participants consenting

An initial response rate to the cohort from the invited participants was 15% for this study of those invited. 125 patients of an invited 845 returned the consent forms. This was broadly in line with and anticipated response rate of 12 – 15% previously observed in two studies using similar GP database recruitment methods. One study recruited to a yoga trial for LBP (12%, 994 of 8638) (Cox et al. 2010) and another recruited to an acupuncture trial for irritable bowel syndrome (15%, 247 of 1651) (Reynolds et al. 2008).

Participant returning baseline questionnaires:

Following consent, 88 of the 125 consented patients returned their baseline questionnaire, providing a response rate of 70% of those who had agreed to be in the study.

A 30% loss of potential participants at this stage of the study was disappointing, but this was an intrinsic part of what the design was planned to advise. Having the consenting process and the baseline questionnaire completion as separate mail outs would have allowed for drop out prior to any substantial work being completed by the researchers or participants and thus may be preferable in a future trial. Attrition within trials is poorly reported (Dumville et al. 2006) and previous trials reporting attrition in their trial generally report it at each data collection period, there is no data available to guide the figures of attrition rate between the consent process and completion of the initial baseline questionnaire. The breakdown of figures for this early dropout rate information is currently unavailable from other studies and maybe useful data to collect to inform future study planning.

Due to the limited resources of this PhD, participants not returning their baseline questionnaires were not followed up or entered into the cohort. While with greater resources this may be a consideration, one of the key

elements considered for this study was attrition. Having a multi-layer administrational process for entry to the study may inadvertently cause those who would have left the study at a later date, to leave earlier, due to the effort required to complete the questionnaires. It could be considered that chasing participants to enter the study at this early stage may not actually be beneficial to a study in the long term, in regard of attrition and their long term commitment to the study.

Participants returning three monthly questionnaires:

At the three-month time point 100% of the 87 participants who were sent their three-month questionnaires, returned their questionnaires fully completed. The 100% response rate is suggestive that the study design, methods and paperwork were appropriate for achieving a reduction in attrition. This further validates the previous point with the early multi-layered recruitment producing early attrition in the pre-randomisation phase of the pilot and the participants remaining showing long-term commitment.

In this nested study design, this is the time point at which the cohort participants became eligible to enter the RCT, thus providing the run in period within the recruitment process. Potentially making it a very valid design methodology for the recruitment of conforming, committed participants.

Recruitment rate of participants from the cohort to the nested RCT

A 65% recruitment rate from the cohort to the treatment trial was achieved for this pilot, which demonstrates the nested design as being a highly effective recruitment model for this RCT. However any additional increase in potential cohort recruitment was significantly limited in this study by the number of participants who scored less than four on the RMDQ. Rendering them ineligible for the nested treatment trial. The 30% loss of potential participants at this later stage was greater than anticipated, regression to the mean would have occurred demonstrating significant

levels of individuals with mild LBP as is evident in society as discussed in chapter one.

The intention of this cohort design was to provide a pool of potentially eligible participants. It is surprising such a significant proportion (30%) failed to achieve the entry requirement of the nested RCT. It may be a consideration to have a minimum entry requirement of RMDQ baseline score to the cohort study in addition to the nested RCT. This would assist in preventing a repeat of such a high total of ineligible participants within the cohort study. Alternatively lowering the eligibility level to be inclusive of participants suffering with low levels of LBP may need to be considered, though this would reduce the value for money of improvement as the participants would have less pain. Thus a separate study of mild LBP may also be indicated.

It is of interest that the participants ineligible for the nested RCT demonstrated a significant proportion of the study population (30%). These participants considered their LBP significant enough to enter a research study but only scored between 0 – 3 on the outcome measurement tool (RMDQ). The mean RMDQ score for those who remained in the cohort for the duration of the study was 1.8. Were this proportion of ineligible individuals reduced through a redesign, the sample size would not require such over inflation (see section 5.6).

This study uncovered significant numbers of participants with low levels of LBP. Further investigation to ascertain knowledge of these individuals would be useful. Are members of society functioning with low level LBP that impact upon them? Or is the RMDQ not sensitive enough to measure and report the significance of LBP in its totality? Does LBP have a greater impact on everyday life than we currently understand or are able to measure? These are all areas that would aid knowledge of LBP as a condition and facilitate research into the area. With additional knowledge using a comparable research design methodology for a future study may provide solutions to these questions. It may be possible in a future planned study to have an additional RCT within the cohort to investigate

characteristics and appropriate interventions of those with low levels of LBP.

The study's conservative sample size target of 16 - 20 participants per group (total sample size 64 - 80) was not achieved. This was primarily due to the significant loss of participants prior to randomisation from the cohort study (who were ineligible for entry into the trial due to a score below four on the RMDQ), so limiting the potential pool of eligible participants from the cohort and the trial numbers.

6.3 Consent and acceptance rate of interventions

124 /125 (99%) participants consented to participate in both the cohort study and nested RCT. Demonstrating a high consent and acceptance rate to the study.

Of the 124 patients that expressed an interest in being offered one of the trial treatments, only three people expressly stated they would not consider one of acupuncture or manual therapy, indicating a high level of acceptability of the treatments (97%).

The three participants who were selective regarding their treatments were still able to be randomised and remain in the study, they were just randomised to either usual care or acupuncture, or usual care or manual therapy, depending upon which they consented to. As previously stated, these participants were excluded from any comparator analyses between acupuncture and manual therapy (alone or in combination).

All participants allocated to an intervention completed their course of treatment, delivering a 100% attendance rate and demonstrating a very high level of treatment acceptance. It would be of interest to assess if a high attendance rate would be repeated with treatments based within an NHS physiotherapy department. It may be possible attendance was higher than expected due to participants holding a high valued opinion of the private practice locally and / or knew and valued the cost of the treatment

they were receiving. Delivery occurring within a private physiotherapy practice setting may have impacted attendance (with convenient access, free parking etc.) or it may have been the design methodology of simply nurturing attuned participants. Further consideration and investigation should be afforded to this area.

The physiotherapists were able to deliver the appropriate interventions to the participants and were able to combine acupuncture with manual therapy for the combined intervention arm for 100% of the sessions required. Demonstrating the ability of physiotherapists to be able to combine the interventions safely and deliver them. This has the potential to save the NHS time and money through the provision of combined interventions in the future.

6.4 Attrition rates, pre and post participant randomisation

One objective of this cohort study with a nested RCT design was to try to limit attrition and resentful demoralisation as far as possible in the early stages of the study. Through the multiple consenting layers and the run in period of the cohort (described in chapter four) occurring prior to randomisation to the treatment trial, it was anticipated the majority of attrition would occur prior to participants entering the nested trial.

As previously discussed in Chapter three (3.3.4) attrition rates are poorly reported in RCTs. Some reports have suggested that attrition can be estimated between 4 – 28% (Hewitt et al. 2010) and 7 – 67% (Dumville et al. 2006; Leon et al. 2006) with the majority of attrition within an RCT suggested to occur during the first of the follow-up periods (MRC/BHF, 2002; Klaber-Moffett et al. 2006). Therefore, due to the three-month 'run-in' period of this cohort study it was anticipated that attrition subsequent to randomisation in this trial would be minimal.

The initial up take and consent to the cohort study was very favourable at 15% of the GP mail-out. The second mail-out of the initial baseline

questionnaire saw a 30% attrition rate to the study, which was substantial at this stage.

Initially participants considered a letter, information sheet and signed a consent form, a relatively modest activity level. However on receiving the more arduous task of completing the initial baseline questionnaire 12 page booklet, some may have considered this to be extensive or did not have the time to commit to its completion. It may be that completion of the initial baseline questionnaire would be more appropriately completed at the time of consent, so potential participants were aware of the level of input required and able to make an informed decision. It may be that the 12-page case report form questionnaire is too extensive and would need consolidating for a full-scale study.

Though a sizeable attrition rate was seen initially, the attrition occurred before the participant or researcher on the study completed any 'substantial work'. This could be considered a positive time for participants to leave the study if they were likely to drop out subsequently at a later stage. However, such extensive attrition was unusual at this stage, and did reduce the number of available participants quite considerably for this pilot study. If the trial design were to remain unchanged an inflation of any future sample size would need to be considered to account for this stage of attrition.

The dual mail-out consenting process was an unplanned change to this study, introduced as a condition of the 'York and Humber Ethics Committee Review Board', at the ethical review of the study. The committee considered it potentially confusing for participants to receive information at the initial contact point regarding the cohort study and the nested RCT. They advised separate information sheets and mail-outs; which were hence incorporated into the induction process.

It may have been that this dual consent process led to early attrition due to participant confusion and possibly a lack of managing the patients' expectations. Alternatively the additional information provided to them

with the second information sheet and questionnaire may have allowed them to make a more informed decision on their non-participation. No data or information from the participants was collected on this, to gain feedback for participant reasons for not wishing to participate in the study. Further investigation into early participant attrition may be useful knowledge to gain and to consider the opportunity for feedback being designed into the methods of a future trial. The process of consent for this study would require redesign and consideration to ensure consented patients are fully informed of the study requirements and expectations, prior to launching a full-scale study.

Attrition between the second consent phase and randomisation for the RCT was extremely low at 1% for this study, with only one patient withdrawing from the trial completely and one patient requesting to remain in the cohort. The additional level of the consenting process may have increased early attrition or perhaps brought attrition that would have occurred later forwards within the process. If it were the later, this provides valuable information into the necessity of a cohort study or run in period being used within studies to minimise participant attrition. An alternative future trial design methodology maybe to consider the incorporation of additional levels or separating paperwork into separate mail-outs to function in the facilitation of early attrition prior to actual recruitment or randomisation. Such a methodology may reduce the need or additional cost and time of using a nested cohort study or a trial run in period.

6.4.1 Zero attrition:

A low (in this case zero) attrition rate post randomisation is a significant and aspirational result for this study. Achieving such low attrition within research studies maybe unrealistic on a larger scale, however working towards it would provide more reliable results, improve the validity of trials and reduce bias within studies. It would also significantly reduce the overall cost of research, aiding funding for future studies.

Zero attrition and 100% response rate of paperwork (with 95% completion rate) and attendance post randomisation demonstrated the effective methodology and appropriate design for this study.

Such a low attrition rate post randomisation was not predicted and provides useful information on a desirable and valid design for future investigations and compares extremely favourably with other back pain trials. For example, three trials (UKBEAM, a CBT trial for low back pain, and a trial of yoga for low back pain) had attrition rates of 25%, 22% and 13% respectively (UKBEAM, 2004; Lamb et al. 2010; Tilbrook et al. 2011); all of these exceed the upper 95% confidence interval limit of 6% for attrition (Dascanio et al. 2014).

6.5 Baseline characteristics of the participants

Differences were seen in baseline characteristics of the participants in this pilot study, which would not be expected in a larger full-scale trial.

Identifying such a difference in female to male ratio between the groups was unusual following randomisation, however such imbalances are not unexpected with a small sample size. There were a significantly greater number of females than males across the whole study and within each trial arm. It is documented that females are more likely to enter trials than their male counterparts (Kennedy-Martin et al. 2015). This may have contributed to the overall higher proportion of women in this pilot.

The age demographic of the pilot study had a mean score of 46.4 in the study, which appear to be in-line with the age demographic of individuals who suffer with LBP and those who enter trials (Nordeman et al. 2016; Kennedy-Martin et al. 2015).

6.5.1 Baseline characteristics for RMDQ and MODI

The baseline score for the cohort participants was significantly lower than that of the nested RCT for both the RMDQ and the MODI. This was primarily due to the minimal entry score of four for the nested RCT. It was not anticipated that a minimal entry score for the cohort study would be a requirement, as participants who scored zero; one or two on the RMDQ were not expected to wish to enrol in a study of LBP. However many participants scoring less than four enrolled in the study and further consideration needs to be given as to why this occurred.

The RMDQ measure records a participant's symptoms on the day they complete the questionnaire and it may be for some that on that day their back was less symptomatic, or it may be that the RMDQ may not have been sensitive enough to measure what the participants consider LBP symptoms that are significant to them. These need exploring further.

A full-scale study may wish to incorporate a minimum entry level score of participant to the cohort study, or may wish to collect information upon this group of individuals who function with low levels of LBP.

It was observed that those allocated to the usual care group recorded the highest RMDQ scores compared to those who had been randomly allocated to the active treatment groups. It was also observed the combined intervention group had the lowest scores on the RMDQ and the MODI at 7.0 and 19.2 respectively at baseline. These between group variations were random and baseline scores were adjusted for in the analysis, therefore the anomalies observed would not have impacted upon the study results.

6.6 Exploratory regression analysis

6.6.1 Roland Morris Disability Questionnaire (RMDQ)

Analysis of the RMDQ scores showed an improvement in all trial arms. This might reflect regression to the mean, the placebo effect and/or the natural history LBP and not necessarily the effectiveness of any treatment modality. The analysis suggested that the manual therapy alone group intervention showed greater improvement in symptoms compared to the acupuncture alone group and the combined intervention group for the RMDQ, though statistical significance was not achieved.

This pilot study was not powered to detect any interaction or additive effects, which may have been occurring within the combined treatment group of this pilot study. A future study is not planned to assess for interaction and / or additive effects, however further consideration for interpreting interaction effects or additive effects via a regression / ANCOVA analysis would need to be contemplated for planning a full-scale well-powered trial. The indication from the pilot is suggestive of a limited benefit in combining the two interventions studied, however given the small sample size this needs to be interpreted with caution, the pilot was designed to investigate the practicality of combining the interventions within the study and not powered to assess efficacy.

As previously discussed, (section 4.5) when the RMDQ was developed, Roland and Morris (1983) noted that one of the difficulties in conducting LBP research, was a lack of suitable outcome measure, and thus they developed one. They developed the RMDQ questionnaire to be short, simple, sensitive, and reliable, advocated for its use in future trials. It has become one of the most commonly used LBP measures in research (Longo et al. 2010).

However, in conditions such as low back pain, patients' symptoms vary daily and weekly; a participant could have had seven days of severe pain followed by less painful days. One of the limitations of the RMDQ is that

patients complete the questionnaire based on how they feel on that day only. With varying conditions of LBP, this may not give an accurate reading of a participant's standard level of back pain over time. These concerns should be consistent across all the randomised groups limiting any bias; however it remains a limitation of the questionnaire, regarding entry level to trials, adequacy of scoring and accuracy of scores between the study time-points.

6.6.2 Modified Oswestry Disability Index

The regression analysis of the MODI scores showed an improvement across all trial arms, with the manual therapy arm (seeing the greatest improvement) and the acupuncture arm achieving improvement over usual care. This was not the case in the combined intervention arm, with no improvement over usual care.

The MODI showed a greater difference in participants in the manual therapy arm, however statistical significance was not achieved, and the pilot study was not powered to detect this.

6.6.3 A comparison and analysis of the RMDQ and the MODI outcome measurement tools

There is a requirement for the outcome measures for any trial to be valid, reliable and sensitive and both these outcome measures have been shown to achieve this within research studies (Morris et al. 2015).

The participants completed both outcome measures in the pilot and appeared happy to complete them (97% completion rates), suggesting it is possible for two outcome measures to be used within a trial. The use of two outcome measures may give greater scope to a trial, a primary measure would need to be identified however a second measure may aid data collection (Streiner and Norman, 2008).

While the observations are not inferential in this pilot, both outcome measures appeared to perform consistently to each other with no significant differences between them, improvements were seen in all trial arms by both measures. The results show some sensitivity to change for both measures, but it is not large enough to determine which measure should be used over the other. Using both measures for a full-scale trial analysis with a greater sample size would be useful in understanding the differences between these measures. This would be supported by a recent systematic review which concluded there was no defining reason to choose one of these measures above the other, they both were valid and reliable and further high quality research was required to determine any superiority (Chiarotto et al. 2016).

Another study compared the outcome measure results of 11 LBP trials (Morris et al. 2015). Each trial used at least two of the seven identified back pain measures in their studies, highlighting the use of two outcome measures as a norm. The study concluded that data gained from different outcome measures should not be pooled for meta-analysis due to the lack of agreement and differing results between the measures (Morris et al. 2015). This has implications for the way trial results can be interpreted via systematic reviews and pooling of data for meta-analyses. The use of two outcome measures in all trials may assist in pooling data across studies.

It would be useful for the research community to investigate and agree upon a small set of appropriate outcome measurement tools for LBP. Standardising outcome measurement tools would allow for meaningful comparisons and a greater capacity of pooling data. However if the evidence base did not support the convergence of practice and consensus, then further research may be required before an agreement can be achieved (Skrabanek, 2003).

While I had planned to compare the two outcome measure (RMQD and MODI) to assess which would be most appropriate tool for a full-scale trial, insufficient data was available from this pilot study. Any observations seen and described (section 5.4.4 and 5.4.5) were from a very

small sample of participants and thus the small sample size prevented any meaningful analysis. Specific criteria to compare and analyses the outcome measures in a future trial would need to be planned and specified in advance of conducting a full-scale study.

Future research should give consideration to the ability and sensitivity of the individual outcome measures to assess the condition being analysed and also the interventions being studied. While extensive research has been conducted into the performance of outcome measures, the research of outcome measures predominantly focuses upon the patient's experiences and the condition being studied. Differences may exist in the sensitivity of the measures, which make some more suitable or not for use with complex interventions. Further research in this area would add to the research knowledge base.

6.7 Ancillary data

Completion rate of the SF12, EQ5D, VAS scale and resource use questionnaires were good at 85% but were not as high as the RMDQ and MODI. Further consideration would need to be given to the length of the 12 page questionnaire booklets and if further guidance may be required for this to be improved in a future study.

In the resource use section of the questionnaire some inaccurate data was identified, with some individuals indicating the trial intervention sessions they had received on the forms. Future use of this questionnaire would require further instruction to participants and clarification not to include their trial intervention sessions on the resource use forms thus preventing skewing of data. While these individuals were contacted by telephone where possible, due to limited resources further follow up was not possible, any future study should plan for additional time and resources to follow up participants to collect missing data and check any potential inaccurate data inputted. Additionally in the "other sections" on the resource use questionnaire, it was repeatedly left blank, if participants did

not have any information to input, further participant instruction or the option to select 'none for these' would considerably reduce the missing data in these areas. Alternatively the research team may discuss and agree a plan for recording uncompleted 'other' sections in the resource use i.e. a box left blank indicating no use of 'other' resources.

A future study would need to have a comprehensive plan for addressing missing data and uncompleted sections of the questionnaires including the resource use section, time and resources to follow up any missing or questionable data.

6.8 Sample size calculation for a full trial

Recruitment from the cohort to the nested RCT and the potential of participant ineligibility would need to be considered and planned for in a sample size calculation of a full-scale trial. Within this pilot 87 participants were recruited to the cohort study, but only 57 were recruited to the nested RCT due to ineligibility of 28 participants and 4 participants not consenting. This resulted in the study objective of recruiting 16 participants to each group not being achieved, however the recommended minimum of 12 for a pilot study (Julious, 2005) was exceeded in all groups.

Using the factorial trial design with analysis "at the margins" means that evidence of efficacy can be obtained from fewer participants than would be required if acupuncture and manual therapy were tested individually in two independent trials (Finlay et al. 2003). However, in a full-scale trial a greater number of participants would be required to investigate the effectiveness of interventions proposed and even more so to test for additive effects and interactions between the interventions (see section 5.5.2.1).

Trials are notorious for not achieving their recruitment objectives with many failing to reach their recruitment targets (Huang et al. 2018), for

example; a study comparing acupuncture with conventional therapy, conducted and reported their RCT with 186 participants, while reporting that a sample size of 380 was required to reach sufficient power (Molsberger et al. 2002). They did not achieve their recruitment aim; thus the study's statistical power and reliability were reduced. The failure to reach planned recruitment is substantial amongst trials, and has significant implications to the scientific validity of trial results, as well as consequence to the financial, ethical, policy makers and stakeholders of trials.

A future full-scale trial of this pilot would need to ensure the recruitment of sufficient number of GP surgeries and recruitment outside the database recruitment model to ensure individuals who did not report to their GP also had the opportunity to partake in the study. A comprehensive plan would be needed to ensure the sample size recruitment objective was achieved for a future full-scale trial of this pilot.

6.8.1 Appropriateness of the design for evaluating a population with LBP

The cohort model proved to be an efficient method of recruitment to the nested factorial trial, with no post randomization attrition rate in this pilot study. The information from this pilot trial provides a useful knowledge for informing a full-scale design and is evidence for further use of this design methodology, as it provided a significantly higher rate of recruitment comparative to recruitment directly into similar trials.

Particularly notable is the compliance of the participants recruited to the nested RCT from the cohort study. As discussed, this pilot achieved a high response rate and commitment following randomisation. Indicating its use for the recruitment of participants who are compliant for future studies. Accomplishing this across RCTs would provide greater accuracy and validity of trial results and reduce the risk of errors.

The study methodology conducted for this pilot differed from original nested cohort design proposed by Relton et al. (2010). In the original

Relton design participants are not specifically told about the possibility of future treatment options available. The difficulty of this approach was that failure to alert the participants at the outset, may mean a refusal to take up the treatment offered later, which might lead to treatment dilution. In this study the possibility of future treatments was presented in the initial stages to avoid this problem (Dascanio et al. 2014).

6.8.2 New findings with this design:

This was the first time this study design was used for a study of LBP and it was seen to be appropriate and effective. The nature of this nested cohort study following participants with remitting and relapsing conditions over a period of time, provided two new opportunities not previously considered for the recruitment to this pilot study. These benefits had not been described by previous research studies using this design (Dascanio et al. 2014).

Firstly, by including the cohort of low symptom patients we could, if the trial had been large enough, have supplemented the randomised analysis by including the cohort participants in a regression discontinuity analysis (Dascanio et al. 2014). This would give the opportunity for an analysis of individuals more generalisable to the population suffering with LBP, inclusive of the whole cohort not just those willing to enter a trial. Further consideration of this analysis would need to be given for a future full-scale trial.

Secondly, using the design for remitting and relapsing chronic LBP conditions, allowed some participants, who initially were not eligible due to low symptom scores, to become eligible later in the trial, and thus could be randomised. In a typical randomised trial design these patients would have been lost from being included in the randomisation and the trial data. The ability to recruit participants from the cohort at varying time points through a trial may prove beneficial to future trial design to maximise recruitment.

In this pilot we were able to offer the opportunity to two of the participants at their six-month time point, both subsequently joined the trial and were randomised for entry into the RCT at six-months. The time constraints of this pilot study prevented further entry into the trial following the recruitment deadline and thus the further nine participants who became eligible to enter the study at their six-month time point later in the study were not able to be admitted.

Considering this trial design and how the time continuum of recruitment could be developed further to include those participants with fluctuating conditions should be a priority for research studies. There is also random variation that occurs on a daily basis of the measurements from the outcome measures tools, thus participants may be missed. A possible longer term cohort monitoring of the condition prior to the trial to gain additional understanding of the nature of the condition would be useful, followed by several recruitment time points over an extended time period would be beneficial to develop knowledge in this area. An extended follow up may also provide more informed evidence into how to effectively manage LBP as a condition in the longer term.

The nested factorial RCT combined acupuncture with manual therapy and this was shown to be an effective way of analysing the interventions, assessed on a larger scale, effects of interactions and additive effects would also be able to be seen with this model.

6.9 Strengthens and weaknesses of the pilot study

The study design was shown to be achievable and effective, however if it were to be scaled up there would be additional work and cost for the researchers to follow-up the non-randomised cohort. It is possible that this would not be a cost-effective use of research resources unless information from the non-randomised group was used. A value of information analysis could be used to explore this and inform future decision-making (Claxton, Sculpher and Drummond, 2002). The non-randomised cohort may

improve recruitment for the study of LBP, as some patients may become eligible who previously were not. As mentioned in 6.7.1, in a larger study the trial based analysis can be supplemented with a regression discontinuity analysis, which would improve study inference. However, arguably the resources spent to obtain these benefits may be better used to increase the overall sample size of the participants who could be randomised. Consequently, it might be more cost effective to modify the design to not follow up the ineligible participants (Dascanio et al. 2014). Analysis can be planned to inform the cost benefit of this and aid the design of an efficient study (Claxton et al. 2002).

The limitations of this study mainly occur due to the limited sample size and the lack of ability to achieve statistical significance in the data collection. However, this was a pilot study that never planned to be powered to detect a difference between the trial groups and thus the results were exploratory to inform a full-scale study (Dascanio et al. 2014).

Another significant weakness of the pilot study was the limited process evaluation conducted for the study. No detailed analysis of the recruitment processes, or the fidelity of the participant or therapist journeys of involvement in the study was conducted. The pilot study did not consider or provide detailed analysis of the implementation of the trial interventions, if any causal mechanisms were occurring, or if any contextual factors were at play during the pilot study (see section 8.3 for further discussion). A future full-scale study would need to incorporate all of these factors into its analysis, providing a 'study of a study' approach (Moore et al. 2015; MRC, 2008; Craig et al. 2008) to generate knowledge, help interpret the study and inform implementation and future research.

6.10 Implications

This pilot study of a cohort design with a nested, factorial RCT, provided useful information to inform a future study design for a full-scale RCT.

Several aspects from this pilot require further consideration when designing a full-scale study, as outlined below.

Further thought would need to be given as to how to access those potential participants in the population who live with the relapsing and remitting condition of LBP but do not access their GP. Investigating the reasons why they choose not to report their condition, i.e. possibly due to previous experiences, mismanagement of their condition in the past, lack of support, access to direct treatment, self-management; this would be useful knowledge gain for future research in this field.

Further consideration would need to be given regarding a minimum entry level on the RMDQ for entry to the cohort study, aiming to limit the number of potentially ineligible trial participants. However as discussed this would have excluded those participants who became eligible at a later time point. An alternative may be to lower the entry-level score to the RCT. Four or more on the RMDQ maybe too high for many individuals experiencing LBP, as demonstrated within the pilot.

Discussion of the need for a cohort model is also required. If the cohort is primarily to be used as a way to reduce attrition, then the inclusion of several layers of paperwork may be sufficient to achieve the early attrition within the trial. If the cohort model were to be used as a recruitment tool to the trial with participants entering the trial at different time points throughout the trial, then it would be a useful to include the cohort as part of a future design. Though adequate consideration would need to be given to the cost and additional workload it would add to a trial.

Using this research design in large treatment trials of interventions for musculoskeletal conditions would be recommended from this pilot study (Dascanio et al. 2014). The most appropriate methodological design should be considered e.g. a factorial trial design, for investigating complex interventions. The comparison of treatment interventions with one another and in combination in research studies may be an opportunity to find advancement in the management of LBP. This study design has the

potential to aid the establishment of preferential treatments and guide any additional benefits of offering additional or combined treatments for LBP (Gleiss, 2017).

7 Systematic Review: Acupuncture Versus Manual Therapy for LBP

7.1 Introduction

Thus far I have discussed the impact of LBP upon society and reviewed the international guidelines and the NICE guidelines in the UK, identifying the need for further RCTs. I have identified limitations and bias within RCT designs and proposed a pilot cohort with nested factorial RCT study design for a study of acupuncture and manual therapy. I have conducted the study, informed by a review of literature undertaken in March 2010. The results of the pilot study have been presented and discussed in chapter five and six of this thesis.

In this chapter I will present a systematic review and meta-analysis conducted in July 2017, which builds on and extends the review of literature in chapter two. Following the results of the pilot study presented in chapter five and discussed in chapter six, there is an indication for further study in this area. Prior to conducting a full-scale study, undertaking a systematic review to provide the most current evidence base available to inform a trial is indicated and would be considered best practice.

7.2 Indication for a systematic review

A lot of research is being published and systematic reviews are fundamental tools for aiding clinicians, health care providers and policy makers to remain up-to-date and implement best practice. Systematic reviews allow evidence to be summarised accurately and reliably and are more easily accessible to clinicians and policy makers in times of limited resource (Grimshaw et al. 2012; Liberati et al. 2009). Systematic reviews

have been suggested as a vital mechanism for getting research into clinical practice (Carrasco-Labra et al. 2015).

Furlan et al. (2005) suggested that the use of acupuncture as an adjunct to other interventions for chronic LBP should be considered. A recent review of systematic reviews of acupuncture for LBP supported this view stating:

> *"Systematic reviews of variable quality showed that acupuncture, either used in isolation or as an adjunct to conventional therapy, provides short-term improvements in pain and function for chronic LBP. More efforts are needed to improve both internal and external validity of systematic reviews and RCTs in this area."*

(Liu et al. 2015)

A study of acupuncture and manual therapy interventions and their combination for LBP would be a useful addition to the current research knowledge base and perhaps provide some clarity on the efficacy of acupuncture and combining interventions as a method to alleviate the chronic issue of LBP within society.

From previous searches I have conducted, it is clear that RCTs combining acupuncture and manual therapy in a factorial design or other design do not currently exist (other than the pilot study for this thesis). While this lack of evidence is suggestive of the need for a new study alone, it remains important to fully review all literature systematically comparing acupuncture and manual therapy for LBP as a measure to fully inform any future study.

Adding to the knowledge base for the treatment and prevention of LBP would aim to inform improvements in care and a reduction of LBP, potentially saving societies in the UK and internationally. There remains considerable deliberation as to whether acupuncture is effective and cost effective provision for the treatment of LBP. The research currently does not conclusively guide policy makers and is open to interpretation, as was

evident from the variation in clinical practice guidelines discussed in chapter two.

The UK NICE LBP guidelines (2009) recommended acupuncture and manual therapy (amongst other interventions), however the updated NICE LBP guidelines (2016) removed the recommendation of acupuncture citing lack of robust evidence (discussed in chapter two). NICE's decision was in contrast to the American guidelines, who recommended acupuncture and other interventions as first line treatments in favour of medication, to the updated LBP guidelines (Qaseem et al. 2017).

Completing a systematic review comparing acupuncture with manual therapy for the treatment of LBP would provide an understanding of the scope of the current evidence base. Its aim was to investigate the most effective and potentially cost effective method of the two treatments for LBP, understand any uncertainties and gain insight into previous study designs and recommendations. It would also be intended to add the results of the pilot RCT conducted for this thesis into a meta-analysis to view a larger data set to investigate if any further insight can be gained.

In addition to informing a future study, the results of this systematic review would aim to provide a focused conclusion to aid practising manual therapists (i.e. physiotherapists, osteopaths, chiropractors), acupuncturists and health-care decision makers in guiding funding and treatment selection for individuals with LBP (Manheimer, 2012).

The aim of this systematic review was to determine the strength of evidence for a recommendation of acupuncture and manual therapy for LBP for the benefit of patients and policy makers.

7.3 Objectives

The objectives of the systematic review were:

- *To determine the effectiveness of acupuncture versus manual therapy for LBP*
- *To determine the availablity of literature for acupuncture and manual therapy for LBP*

7.4 Methods

This systematic review was conducted based on the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) Statement (Liberati, 2009) to ensure the provision of transparency and completeness, of the reporting of the review and meta-analysis. It also used GRADE (Grading of Recommendations Assessment, Development and evaluation) a transparent framework to develop and present a summary of the evidence, establish the strength of the evidence and to propose a systematic approach to clinical recommendations (Schümenann, 2013; Guyatt et al. 2013; Guyatt et al. 2011).

This systematic review included a PRISMA flow diagram, figure 7.1 and aimed to ensure it was reproducible and provide confidence in the results. The study criteria were established in advance and followed a protocol (appendix E0) (however the SR protocol was not registered with PROSPERO or published prior to commencing the review); the recording of all decisions with explanations was conducted at each stage of the review (appendix E).

7.4.1 PICOS study inclusion and exclusion

The inclusion of published studies was based on the requirement to fulfil the following PICOS criteria for selection:

- *Population: People suffering low back pain*
- *Interventions: Acupuncture and manual therapy*
- *Comparators: Acupuncture vs manual therapy with or*
- *without an additional control group*
- *Outcome: Reduction of pain, improvement in function*
- *Study design: Randomised controlled trials*

(GRADE, 2013. Stone, 2002)

Population:

Age demographic: As per the protocol, studies of adult participants of all genders aged between 18 - 65 years with a diagnosis of 'non-specific' LBP were included, however in a deviation from the protocol the age demographic was expanded (as detailed below) due the revisions of the NICE guidelines.

LBP is less common in people below the age of 18 and the management of children with LBP differs from the adult population, it would be inappropriate to review these two population sets together. However, within the time frames of this thesis the recent NICE LBP guidelines (2016) updated their parameters to consider individuals from the age of 16 years within their guidelines and thus any future planned study would follow the NICE guideline format. Therefore, in a deviation from the protocol for this systematic review, any studies inclusive of an age range from 16 years were considered and included. Studies of those below the age of 16 years were excluded.

Individuals over the age of 65 years are more likely to suffer from serious spinal pathology and co-morbidities that may affect their response to low

back treatment (UKBEAM, 2004) and thus would be inappropriate to consider. The updated NICE LBP guidelines (2016) however, considered individuals above the age of 65 years. Any future planned study would follow the NICE guideline format, therefore in a deviation from the documented protocol for this systematic review, any studies inclusive of individuals above the age of 65 years were considered; conditional to the provision of information outlining their age range and any serious spinal pathology or comorbidities suffered. Studies of those over 65 years were excluded if serious spinal pathology or comorbidities impacted upon the study group or if serious spinal pathology or comorbidities were not reported.

Non-Specific LBP: Studies where a diagnosis of 'non-specific' LBP of the population was determined by a General Practitioner (GP) or other healthcare practitioner (Physiotherapist, Osteopath, Chiropractor, Nurse practitioner) only were included to ensure appropriate screening had been conducted and an accurate diagnosis determined.

> *'Non-specific' Low Back Pain; = a musculoskeletal problem, not attributable to a specific pathology (Milczarek, 2009).*

Specific LBP: Studies where a diagnosis of 'specific' LBP of the population was determined by a General Practitioner (GP) or other healthcare practitioner (Physiotherapist, Osteopath, Chiropractor, Nurse practitioner) were excluded.

> *'Specific' Low Back Pain: = infection, cancer, rheumatic disease, osteoporosis, post-operative or nerve root involvement. (Milczarek, 2009)*

Studies not specifying the type of LBP studied were excluded, unless the reviewers could establish the LBP type from the reported information.

Time frame: Trials investigating any one or more of acute (one to six weeks), sub-acute (six to twelve weeks) and chronic LBP (more than

twelve weeks) (Milczarek, 2009) were included, to be representative of the population of LBP sufferers.

Following these criteria ensured greater external validity of this review than the planned protocol, and enabled the study to be generalisable to a population between the ages of 16 – 65 years or older without comorbidities who suffer with non-specific LBP.

Interventions and Comparators:

RCTs comparing the use of acupuncture with manual therapy for the treatment of LBP were selected for this review.

Both the acupuncture and manual therapy interventions in any selected studies were required to be conducted by a suitably qualified health care professional, trained in their respective field. Each intervention would act as a control to the other, with or without an additional control group.

Acupuncture was restricted to 'real acupuncture' defined as the insertion of an acupuncture needle into specific acupuncture points (WHO, 2002). The style of acupuncture was not limited for this review to ensure completeness of trial information. If more than one type of acupuncture or two acupuncture arms was studied, they were included if an appropriate comparator arm was also apparent.

Studies of acupuncture with non-penetrating needles, acupressure and laser acupuncture were excluded.

Some acupuncture trials may have considered manual therapy / physiotherapy as 'usual care' or as a control arm; RCTs using this design were considered for the review if the intervention included the use of manual therapy.

Manual therapy; incorporates mobilisation, therapeutic massage and manipulation treatments, all these interventions were included under the classification of manual therapy:

- *Mobilisation; Joint and soft tissue movement within normal range*
- *Massage; Manual manipulation or mobilisation of soft tissues*
- *Manipulation; Low amplitude, high velocity movement taking joints beyond normal range*

(NICE guidelines, 2009)

Studies of all these types of manual therapy were included. Studies using mechanical devices to deliver manual therapy or light touch / sham manual therapy techniques were excluded.

Outcomes:

Studies were included if a primary outcome measure focused on 'Pain Intensity', 'Quality of Life', 'Functional Status' or 'Occupational Status'. These are considered to be key areas of focus in the discipline of LBP and are important areas of attention for patients with LBP (Chiarotto et al. 2018; Maughan and Lewis, 2010; Furlan et al. 2008).

Primary outcomes:

- *Quality of Life: e.g. EQ5D, SF-36, SF-12, Patient self-efficacy questionnaire (PSEQ)*
- *Functional status; e.g. Roland Morris disability scale, Oswestry disability index, Quebec back pain disability scale, SF-36, Sickness impact profile, Patient Specific Functional Scale (PSFS)*
- *Occupational status; e.g. Return to work status, number of sick days off work*
- *Pain intensity; would be included if used in combination with one of the above measures e.g. Visual analogue scale (VAS), Numerical pain rating scale, Numerical Rating Scale (NRS), McGill pain inventory*

(Chiarotto et al. 2018; Maughan and Lewis, 2010; Resnik and Dobrzykoski, 2003; Furlan et al. 2008)

The outcome measures are useful tools and their objective is to provide the researcher, clinician and patient with a change in the measurement that parallels with a meaningful clinical change (Hägg, Fritzell and Nordwall, 2003).

Other outcomes were not considered for this review; e.g. economic outcomes, patient satisfaction, adverse reactions, negative consequences of the interventions, side effects, recurrence, fear avoidance behaviours, medication, depression e.g. Hamilton depression rating scale (HAMD). If a primary outcome did not measure quality of life, functional status, pain intensity, or occupational status, they were excluded. Other measures not listed above were only considered if they were appropriate to LBP and evidenced to be reliable, accurate and valid.

While the collection of data on adverse events was not initially planned at protocol stage, when actually conducting the review I did include information on adverse events to ensure completeness and in order to provide the information required for clinical relevance. This addition was a deviation from the protocol.

Study Design:

All randomised control trials that meet the inclusion criteria were included in this review. Only RCTs were selected as they provide the strongest form of evidence, they are the most robust in their design to ensure they control for selection bias, regression to the mean and temporal changes. Randomisation minimises biases in the allocation of participants to each arm of the trial and helps ensure the internal validity is robust (Torgerson and Torgerson, 2008).

RCTs are required to have documented adequate randomisation e.g. computer generated, sealed envelopes, independent allocation. Studies without adequate or documented adequate randomisation were included in the narrative analysis of this systematic review but excluded from any meta-analysis due to the risk of bias in the studies.

Quasi-experiments were not included in this study. They can be inherently biased and are vulnerable to temporal changes, regression to the mean and subversion bias, questioning the internal validity and results of the study (Torgerson and Torgerson, 2008). Inclusion of results with inherent limitations could bias the results of this systematic review. All other forms of study design were also not considered sufficiently rigorous in their design and may be subject to biased results and thus were not considered for this review.

7.5 Language selection

Only studies in the English language were included. Studies with an English language translation were included and considered as part of this review. The potential increased cost implication to this review of non-English studies; in terms of the time, resources and translation services was not possible for this systematic review as part of this PhD project.

It is recognised that the inclusion of English only papers may potentially introduce bias and affect the validity of the review (Moher et al. 1998). However having no language restriction may also lead to concerns over publication bias, especially in the field of acupuncture and complimentary medicine, where a majority of study and evidence is based in Asia; publications in Asian journals have been reported to be of low quality and it is suggested only positive results are published (Ernst, 2016). Vickers et al. (1998) have demonstrated in their review that some countries publish only positive results, they reported no trial published in China or Russia found an investigated treatment to be ineffective. They also reported a very strong tendency towards the publication of positive results (China 99%, Russia 97%, Japan 89%, England 75%). Accessing negative findings from China and Russia may not be possible and would create further increased resource, time and translation costs with potentially limited success.

All papers were carefully considered, adhering to the inclusion criteria to ensure eligibility and quality of included trials. Searches for unpublished material were implemented in an attempt to ensure publication bias did not impact this review.

7.6 Search methods for the identification of studies:

A computerised literature search with English language restrictions including all publications to July 2017 was conducted using the following databases:

- *EBSCOhost included;*
- *AMED - 1985 to July 2017*
- *CINAHL database -1981 to July 2017*
- *CINAHL Plus with full text - 1937 to July 2017*
- *CSP Online Library Catalogue – 1937 to July 2017*
- *MEDLINE – to July 2017*
- *SPORTDiscus – to July 2017*

The search strategy used for EBSCOhost is documented under appendix E1.

- *ProQuest Dialog Healthcare included;*
- *British Nursing Index - 1994 to July 2017*
- *Allied & Complementary Medicine - 1985 to July 2017*
- *DH-DATA: (Health Administration, Medical Toxicology & Environmental Health) 1983 – July 2017*
- *MEDLINE 1946 to July 2017*
- *EMBASE 1947 to July 2017*
- *EMBASE Alert – July 2017*

The search strategy used for ProQuest is documented under appendix E5.

Grey literature was searched through SIGLE (System for information on Grey Literature in Europe) and HSRProj (Health Services Research Projects in Progress).

Other databases searched included:

- *Unpublished databases – System for Information on Grey Literature in Europe (SIGLE)*
- *Unpublished databases - National Research Register; Health Services Research Projects in Progress (HSRProj)*
- *CENTRAL, the Cochrane Library*
- *ACULARS (Acupuncture Literature Analysis and Retrieval System)*
- *Acubriefs.com to July 2017*
- *The Clinical Trials Register and the ISRCTN Registry*

A search strategy was performed following the Cochrane Back Review Group (CBRG) guidelines (Furlan et al. 2015), it was adapted for our search as we were searching for trials with comparator treatments (acupuncture versus manual therapy) and not placebo or drug comparators alone. Specific terms for LBP were used, though terms for conditions considered as 'Specific LBP' were adapted for our search. The search strings for 'EBSCO' and 'ProQuest' are provided (appendices E1, E5) and these search terms were adapted for the other databases as appropriate to ensure the key components, terms and mesh headings were correct for each database.

All reference lists and bibliographies from systematic reviews and all selected original articles were reviewed for any additional studies. All duplicate articles were logged and excluded.

Hand searching was conducted for a small number of journals e.g. Acupuncture in Physiotherapy (previously the Journal of the Acupuncture Association of Chartered Physiotherapists) and publications held by the Chartered Society of Physiotherapy. Hand searching was performed due

to these journals historically not being well covered by electronic databases (appendix E9d).

The above search strategy was implemented to aim to ensure a comprehensive and thorough search. Every effort to find unpublished and papers with negative findings was made in an attempt to reduce the possibility of publication bias affecting the results of this review.

7.7 Data collection and analysis

7.7.1 Selection of Studies

For this review the first reviewer (VCD) generated the electronic search strategies for EBSCOhost, ProQuest Dialog Healthcare and the other databases (appendices E1, E5).

The database searches and searches of other sources were then conducted by the first reviewer. Once the search results were completed, both reviewers conducted the identification of potential studies independently. The two reviewers (VCD and AWG) screened the titles and abstracts of all studies using the previously piloted study eligibility form (appendix E12) and either excluded (reasons for exclusion were documented; appendices E2, E6, E10a,b,c,d) or selected as full text for review.

Selected full texts studies were independently reviewed (VCD and AWG), observing the inclusion and exclusion criteria. Study eligibility forms were independently completed (appendix E12), providing reasons for any exclusion (appendices E3, E7).

Reasons for excluding studies were provided to ensure transparency of the selection process and to limit any bias within the review process. Consensus was used for any discrepancies; and arbitration by a third independent reviewer used to resolve any disagreement.

7.7.2 Data extraction

Both reviewers piloted the data extraction form (appendix E13) to ensure consistency, the extraction was appropriate, no errors occurred, and biases were excluded.

After piloting, the two reviewers independently extracted data from the selected studies. The data extraction incorporated authors, year of publication, language, setting, country, study information, methodology, study population, study interventions, study comparisons, study outcomes, randomisation, blinding, data analysis, data to assess risk of bias, results, attrition and funding sources. Data on adverse events was also collected. The objective of two independent reviewers was to reduce the risk of mistakes, data input errors and any relevant information being missed, reducing the introduction of bias (Edwards et al. 2002).

Data extraction was recorded on data extraction forms to ensure transparency of information, consistency, and reproducibility, consequently reducing any risk of bias in this review. If any discrepancy could not be resolved through discussion arbitration by a third independent reviewer was conducted, and they had the concluding decision.

An attempt to retrieve any missing data was planned, by contacting the study's authors.

If multiple publications of the same study existed, all appropriate information was extracted, but the data was treated as one study and analysed once.

Measurement bias can arise due to differences in outcome measurements. High quality trials provide full descriptions of the criteria for measuring outcomes and reduce the risk of bias. All selected studies were reviewed for their reporting of the measurement outcomes, to assess the quality of the studies (appendix 14).

7.7.3 Assessment of methodological quality

Moher et al. (1998) highlighted how the quality of the design, conduct and analysis of a trial can affect the estimate of the efficacy of the assessed intervention, it is therefore imperative to assess if a study is robust.

The assessment of methodological quality including the risk of bias was assessed for this review using the 12 criteria recommended by the Cochrane Back Review Group (CBRG) (Furlan et al. 2015; Furlan et al. 2009; Bombardier, Esmail and Nachemson, 1997) and considered design, quality of methodology, consistency of results, sufficient data, generalizability and risk of bias. This is considered a comprehensive tool in the field of LBP (appendix E14) and relates to the Cochrane risk of bias tool.

Prior to assessing the selected studies, a pilot process of assessing the criteria was performed by both reviewers independently to identify and address any opportunity for misinterpretation or disagreement. Any disagreements were resolved through discussion and any unresolved discrepancies were resolved through arbitration with a third reviewer, who had the concluding decision. Results were recorded on a risk of bias assessment form for transparency of decisions and guidance notes are provided for each reviewer to ensure consistency of decisions.

The 12 criteria were scored as 'yes', 'no' or 'don't know' and reported with reasons for each decision to demonstrate transparency of the decisions. For this review an RCT was considered at 'low risk of bias' of high quality if it meets criteria 'A' (randomisation), 'B' (allocation concealment), 'C5' (outcome assessor blinding) and a minimum of three other criteria.

Due to the nature of many acupuncture and manual therapy RCTs being pragmatic and blinding of clinicians and patients to treatment intervention being unrealistic in many studies, criteria 'C3' (patient blinding) and criteria 'C4' (clinician blinding) was interpreted as the clinician and patient

not being informed of the outcome of their intervention in relation to the study objectives until after analysis of the whole study.

The two reviewers assessed the methodological quality and risk of bias of the selected studies independently to ensure accuracy, consistency and transparency of the review, reducing any risk of bias. This assessment was conducted to ensure any studies with serious flaws were excluded from any meta-analysis, (e.g. exceptionally high attrition rates, or trial conclusions not supported by the reported statistical results), and also to grade the quality of the trials from low too high to guide the strengthen of the evidence presented (Low quality studies with a high risk of bias fulfilled six or less of the criteria, high quality studies with a low risk of bias fulfilled seven or more criteria) attrition rate was also considered for the risk of bias assessment.

Studies with low risk of bias were included in any pooling or meta-analysis of the results, any studies of low quality and a high risk of bias were considered further before any inclusion or rejection from pooling or meta-analysis, a sensitive analysis may be considered if appropriate (Bland, 2000).

7.7.4 Adequacy of interventions

The adequacy of interventions within the selected studies is a subjective analysis therefore both reviewers agreed on the adequacy in delivery of the intervention for each included study. The reviewers both held extensive knowledge and experience in acupuncture and manual therapy and were well informed to assess the adequacy of an intervention.

Each intervention was judged as adequate, moderate or inadequate for the studies; if any interventions were deemed to be inadequate in their delivery of the intervention the studies were excluded from pooling of the results in a meta-analysis. Adequacy included consideration to the type of treatment, the length of session, the number of treatment sessions, the period of time they were delivered over and the therapist delivering the intervention.

Detailed explanations were provided of the reviewer's views of any studies excluded for inadequate interventions. If studies were considered of moderate adequacy, they were given further consideration in relation to quality and the other parameters of the systematic review to decide if they were appropriate or not for pooling in a meta-analysis, with explanations provided.

7.7.5 Clinical relevance

An assessment of clinical relevance of the studies was performed using an adapted version of the assessment guide for clinical relevance (appendix E15) developed by the Cochrane Back Review Group (Furlan et al. 2015; Furlan et al. 2008).

7.7.6 Data analysis

Descriptive data was used to summarise the main characteristics and conclusions of the studies and these were presented.

The quality of the studies was established as recommended by the CBRG (Furlan et al. 2008); with focus upon study design, quality of methodology, consistency of results, sufficient data, generalisability and low risk of bias, and a meta-analysis was considered for good quality studies. A meta-analysis is regarded as useful tool for a systematic review as it provides a clear picture of the evidence, provides a common effect of the study data by pooling the data, and summarises the results of several studies into one single estimate of treatment effect.

The meta-analysis would consider the interventions comparative to each other to consider any differences within the study results. Sub-group analyses were not anticipated as a requirement for this review.

To perform a meta-analysis of the studies for continuous data, the mean, standard deviation and sample size were required for each trial for analysis to occur. If data from a study were inadequate for analysis, the authors

were contacted to request further information (Singh et al. 2017; Bland, 2000).

For continuous data outcomes, mean difference and standard deviations were presented. Any data presented with alternative measurements was converted into standard deviations for the pooling of the data for meta-analysis. Any dichotomous data present was reported as risk ratios or odds ratios and their 95% confidence intervals. Inverse variance methods (Mantel-Haenszel method) were used for pooling of data where appropriate (Bland, 2000).

The software package RevMan 5.3 was used for the meta-analysis. A common estimate of the mean and standard deviation was used, and data presented in other forms was converted to mean values and standard deviations for each study to provide a common study denominator.

Chi-squared was calculated as:

$$Q = \text{sum of (study estimate} - \text{common estimate / standard error)}^2$$

Heterogeneity between the studies would be assessed using I^2. The I^2 is the percentage of variation across the RCTs that are due to heterogeneity rather than chance (Higgins and Thompson, 2002). I^2 was calculated as:

$$I^2 = (Q - df) / Q$$

If heterogeneity / I^2 were below 50% a meta-analysis would be performed to pool the data using the fixed effects model, as 50% or below would be considered as low to moderate heterogeneity. If heterogeneity fell between 50 - 75% then a meta-analysis would be performed using a random effects model, as this would be considered as high heterogeneity. If heterogeneity rose above 75%, pooling of the results was not recommended as it would be invalid to pool the results into a single summary and a narrative analysis would be provided (Singh, 2017; Gagnier et al. 2012; Bland, 2000).

If any data was inadequate for analysis, the trial was excluded from any pooling of the results and presented descriptively. The extent of attrition

bias and the use of the intention to treat (ITT) analysis to reduce the risk of attrition bias were considered for each trial (Torgerson and Torgerson, 2008). Trials using ITT were included, trials not using ITT may indicate bias and low quality, these trials were considered for quality, and attrition levels were assessed prior to pooling of any data for a meta-analysis.

A sensitivity analysis may be performed if weaker (low quality or very small) studies looked to be influencing the results; this was assessed considering outlying results or substantial differences to other studies. A sensitivity analysis without these studies was an efficient way to consider the influence of quality. An analysis of the stronger evidence may be useful, to see if the results differ, giving an indication of the influence of strength of research. If questionable studies exist in the review, an analysis was performed without them to assess their influence on the results. If any treatments were assessed as inadequate, a sensitivity analysis to investigate the impact of their exclusion was conducted to ensure the reviewer's views had not biased the results.

Outcome measurements were analysed together at their primary outcome measurement time point. If the continuous outcome measures were not measured on the same outcome scale the standardised mean difference (SMD) was used. The weighted mean difference (WMD) would be used to provide a standard unit of measurement for the meta-analysis for pooling data. They are weighted by how informative each study is. Studies would be weighted to reflect their importance, the greater the sample size the greater the weighting of the trial for the meta-analysis (Bland, 2000). Forest plots were presented for the results of meta-analysis conducted.

7.7.7 GRADE methods

The GRADE framework was used for this SR to assess the quality and strength of evidence, and to make recommendations based on the assessment.

The GRADE certainty (strength of evidence) assessment considered risk of bias, inconsistency, indirectness, imprecision and other considerations, detailed below and the results of the certainty assessment influenced the quality rating (table 7.1) (Zhang et al. (19) 2018).

- *For risk of bias: a judgement was made regarding whether the risk of bias in the individual studies was large enough that confidence in the estimated treatment effect was reduced.*
- *For inconsistency: examination of whether the studies showed consistent results was conducted.*
- *For indirectness: consideration was given to whether the studies directly compared the interventions of interest in the population of interest, and outcomes where reported.*
- *For imprecision: the GRADE approach focuses on the 95% confidence intervals (CIs), i.e. confidence in estimates of effect, and this was reviewed for the included studies.*
- *Other considerations: the transparency of reporting of adverse events in the studies was conducted for this assessment.*

For the summary of findings: the quality level of the evidence was determined for this assessment by the GRADE criteria, which has four levels of evidence for its quality rating (table 7.1).

Quality	What it means
High	The authors have a lot of confidence that the true effect is similar to the estimated effect
Moderate	The authors believe that the true effect is probably close to the estimated effect
Low	The true effect might be markedly different from the estimated effect
Very Low	The true effect is probably markedly different from the estimated effect

Table 7.1: GRADE – levels of evidence (adapted Zhang et al. 2019;Guyatt et al. 2013)

For relative importance: the GRADE classification of relative importance of outcomes (table 7.2) (GRADE, 2013; Schünermann et al. 2013). This review was assessed using the GRADE criteria, with consideration to the importance of health benefits and potential harms of the interventions.

GRADE Rating Scale								
1	2	3	4	5	6	7	8	9
Of Limited importance for making a decision (not included in the evidence profile)			Important, but not critical for making a decision (not included in the evidence profile)			Critical, for making a decision (included in evidence profile)		

Table 7.2: GRADE relative importance rating scale (adapted from GRADE, 2013)

Strength of recommendation: a GRADE strength of recommendation was provided with justifications, based on the quality of the evidence assessed following the GRADE framework.

7.8 Search strategy results

The electronic searches of the databases listed in section 7.6 was conducted on the 4th July 2017 (appendices E1, E5, E9), the results of the selected studies from the systematic review searches are detailed in section 7.9 and in the PRISMA flow diagram below in figure 7.1.

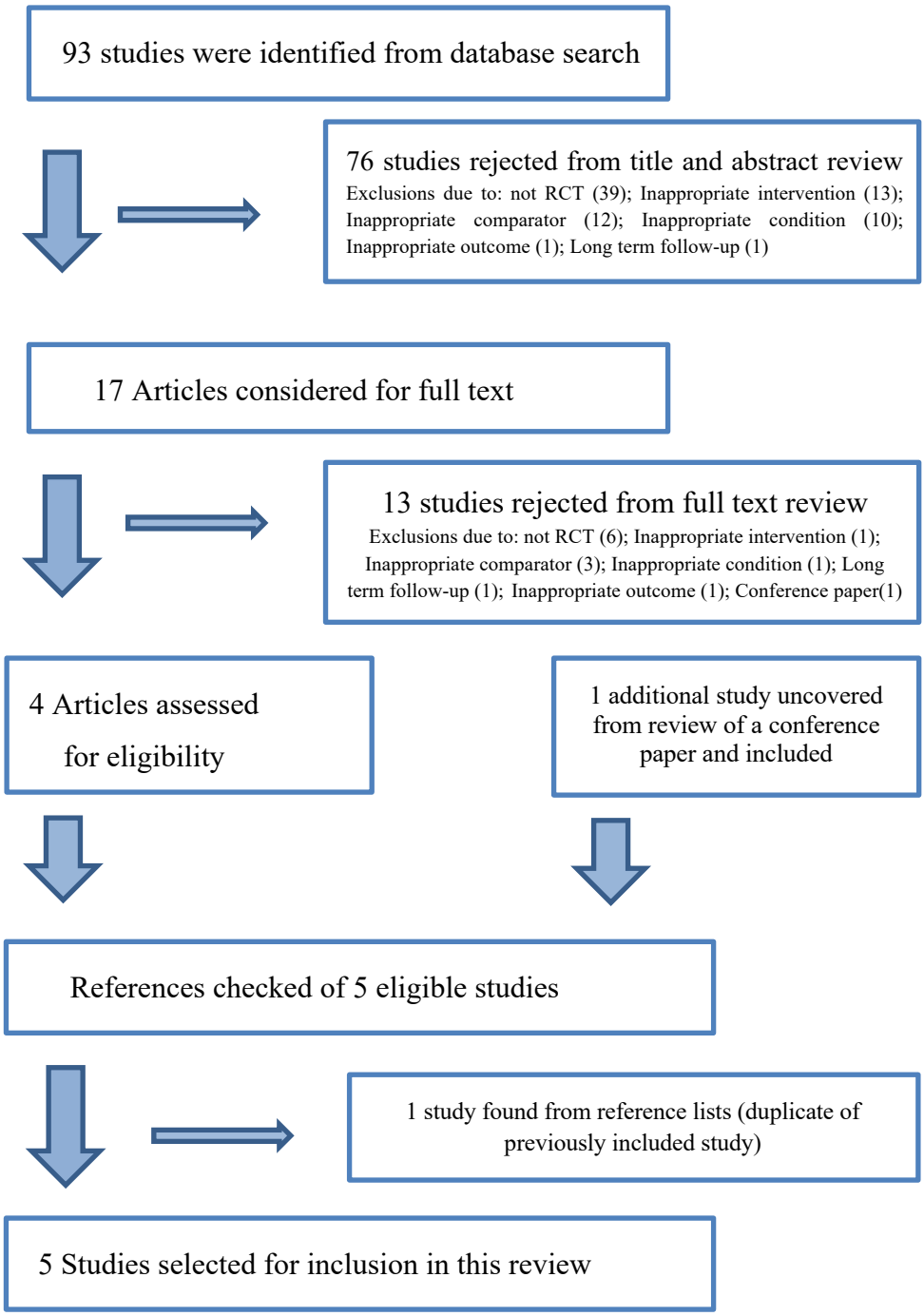

Figure 7.1: PRISMA flow diagram of identified articles via literature search

The results revealed from the individual search engines are detailed in table 7.3 and presented as totals in the PRISMA diagram (figure 7.1).

Search Engine	Search results	Excluded Studies	Included Studies
EBSCOhost	48	45	3
ProQuest	32	31	1
Grey Literature SIGLE	5	5	0
Grey Literature HSRProj	6	6	0
Hand Searching	1	1	0
Reference List Search	1	1	0
Review of Conference paper	-	-	1

Table 7.3: Individual search engine results of database search

As the PRISMA flow diagram presents, of the 93 studies found, the review of titles and abstracts resulted in 76 studies being rejected on the basis of not fitting the parameters of the systematic review: 39 studies were rejected for not being RCTs, 13 studies reviewed different interventions, 12 studies had inappropriate comparators, ten studies had inappropriate conditions, one was a long term follow-up of an included study with duplicate results, one study had an inappropriate outcome with no outcome data provided (appendices E2, E6, E9, E10, E11).

17 studies were selected for full text review (figure 7.1). 13 studies were excluded for not fitting the parameters of the systematic reviews; six studies were excluded for not being RCTs, one had an inappropriate intervention, three inappropriate comparators, one an inappropriate condition, one study had inappropriate outcomes, one study was a long term follow up of an included study with duplicate data, and one study was a conference abstract paper with no reported data (however further follow-up of the conference paper uncovered an additional paper for potential inclusion) (appendices E3, E7, E11).

Four studies (Kumnerddee, 2009; Giles et al. 2003; Cherkin et al. 2001; Giles et al. 1999) were included from the original search, plus one uncovered paper (Dascanio et al. 2014) from a conference study paper, which resulted in five studies remaining for inclusion.

The reference lists of the five studies were checked and one duplicate study was found, and had been included. Therefore five studies matched the criteria for this systematic review and were selected for inclusion (figure 7.1). The narrative break down of each individual search conducted is detailed in appendix E11.

7.9 Selected studies

As the PRISMA flow diagram (figure 7.1) highlights, five studies met the inclusion criteria and were included within this review. No unresolved disputes between the two independent reviewers existed. Table 7.4 details the studies selected for inclusion within this systematic review.

Title, authors and date	Participants /conditions	Intervention
Dascanio, VC. Birks, Y. Torgerson, D. (2014)	Low back pain	Acupuncture, manual therapy and usual care
Kumnerddee, W. (2009)	Back pain	Massage Acupuncture
Giles, LGF. Muller, R. (2003)	Chronic spinal pain	Medication, acupuncture, and spinal manipulation
Cherkin, DC. Eisenberg, D. Sherman, KJ. Barlow, W. Kaptchuk, TJ. Street, J. Deyo, RA. (2001)	Chronic low back pain	Acupuncture, Therapeutic massage and self-care
Giles, LGF. Muller, R. (1999)	Chronic spinal pain syndromes	Acupuncture, a non-steroidal anti-inflammatory drug, and spinal manipulation

Table 7.4: Selected studies included within the systematic review

7.10 Description of included studies

This systematic review included five randomised controlled trials, detailed in table 7.5; two were pilot trials (Dascanio et al. 2014 – (the study conducted for this thesis); Giles et al. 1999), two were full-scale trials (Cherkin et al. 2001. Giles et al. 2003) and one was a small trial with preliminary findings reported (Kumnerddee, 2009).

All the studies were published in English language between the years of 1999 and 2014. 531 participants were included across the five studies. All studies clinically assessed each participant as having low back pain, four with a sub-acute (six to twelve weeks) or chronic (more than twelve weeks) nature and one (Kumnerddee, 2009) with acute (zero to six weeks) or sub-acute (six to twelve weeks) nature.

One of the studies (Giles et al. 2003) and one pilot (Giles et al. 1999) considered spinal pain, however separate results for data sets on back pain and neck pain were provided, to enable the selection of the back pain data only, for the purpose of this systematic review.

All studies compared the effectiveness of the two identified interventions (acupuncture and manual therapy) against each other; Kumnerddee (2009) compared Chinese acupuncture against Thai traditional massage. Giles et al. (2003) trial and Giles et al. (1999) pilot trial compared acupuncture with spinal manipulation, Cherkin et al. (2001) trial compared Chinese acupuncture with massage and Dascanio et al. (2014) pilot trial compared Western acupuncture with manual therapy.

In addition to the identified interventions; Giles et al. (2003, 1999) also included a medication arm; Cherkin et al. (2001) included a self-care arm and Dascanio et al. (2014) had a usual care arm and an arm combining acupuncture with manual therapy. Kumnerddee (2009) did not have an additional arm. All studies reported on pain or function as the main primary outcome.

Study	No of part -icipants	Pain Duration	Interventions	Outcome Measurement	No. in each trial arm & Lost to follow-up	Timing of Primary Outcome & Analysis	Results	Trial Conclusions
Dascanio *et al.* (2014) UK Pilot trial	59	6 weeks or more (Sub-acute - chronic)	Usual care 10 x Acupuncture 30minutes 10 x Manual Therapy 30minutes 10 x Combined A and MT 45minutes Weekly	Roland Morris Disability Questionnaire Modified Oswestry Disability Index	Usual care - 14 Acupuncture - 13 Manual Therapy - 13 Combined – 12 (0 lost)	12 weeks (3 months) (after initial treatment) Analysis of Covariance	RMDQ: Mean (SD) Usual care 9.5 (6.3) Acupuncture 6.8 (4.5) Manual therapy 4.6 (4.0) Combined 5.4 (4.8)	Combined acupuncture and manual therapy had the greatest effect. Statistically significance not achieved
Kumnerd dee (2009) Thailand	18	3 months or less (acute – sub-acute)	5 x Acupuncture 5 x Thai Massage 1 hour Every 2 days	Thai McGill pain questionnaire Visual Analogue Scale (VAS)	Acupuncture - 9 Thai Massage - 8 (1 lost)	10 days (after initial treatment) Paired T-test	McGill: Mean (SD) Acupuncture 2.11 (2.21) Thai Massage (10.25 (11.02)	Significant difference shown for acupuncture post treatment Improvement across all other time points, not statistically significant
Giles *et al.* (2003) Australia	115	6 weeks or more (Sub-acute - chronic)	Manipulation 20minutes Acupuncture 20minutes Medication 2 x Weekly	Oswestry Disability Questionnaire VAS	Manipulation – 35 (2 lost) Acupuncture – 34 (2 lost) Medication - 40	9/52 (after initial treatment) Wilcoxon signed-rank test	ODI: Mean (interquartile ranges) Manipulation 12 (0-29) Acupuncture 24 (11-36) Medication 32 (11-46)	Manipulation had greater short-term effect than acupuncture. The data does not support single use intervention.
Cherkin *et al.* (2001) USA	262	6 weeks or more (Sub-acute - chronic)	10 x Acupuncture 10 x Massage Self-care Weekly	Modified Roland Disability Scale	Acupuncture - 94 Massage - 78 Self-care – 90 (5% (13) lost, no between group figures provided)	10 weeks (after initial treatment) ANCOVA	RDS Mean (95% confidence intervals) Acupuncture 7.9 (6.5-9.3) Massage 6.3 (5.1-7.5) Self-care 8.8 (7.4-10.2)	Massage effective for LBP Acupuncture relatively ineffective
Giles et al (1999) Australia Pilot trial	77	13 weeks or more (chronic)	Manipulation 20 minutes Acupuncture 20 minutes Nonsteroidal anti-inflammatory drug 6 treatments in 4 weeks	Oswestry Disability Questionnaire VAS	Manipulation – 36 Acupuncture – 20 Medication -21 (49 lost, above no remaining, no between group figures provided	4/52 (after initial treatment) Post intervention minus pre intervention measurements	ODI: Median (interquartile ranges) Manipulation 28 (14.5-41.5) Acupuncture 24 (18.5-35.5) Medication 20 (14.5-39.5)	Spinal manipulation superior to acupuncture and medication Statistical significance achieved for the manipulation group results only

Table 7.5: Summary of included studies (RMDQ = Roland Morris Disability Questionnaire, ODI = Oswestry Back Pain Disability Index, RDS = Roland Disability Score)

7.11 Study summary details

Dascanio et al. (2014): piloted a cohort study with a nested RCT. 59 participants aged 18 – 65 with a GP diagnosis of LBP were randomised to one of four groups. Randomisation was conducted via an SPSS computer programme and ITT was adopted (14 Acupuncture, 16 Manual Therapy, 13 Acupuncture combined with Manual therapy, 16 Usual GP care). The baseline characteristics and outcome data are presented in table 7.6.

Parameters	Manual Therapy N = 16	Acupuncture N = 14	Usual-care N = 16	Combined N = 13
Age (years) (SD)	43.9 (13.7)	45.6 (11.9)	43.8 (11.7)	50.1 (9.3)
Male (%)	56	29	31	38
Onset (years)	-	-	-	-
Roland Morris at baseline (SD)	8.0 (4.4)	8.8 (4.3)	11.4 (5.3)	7.0 (2.6)
Modified Oswestry Score (SD)	24.0 (13.6)	29.6 (12.2)	29.5 (15.4)	19.2 (8.0)
Roland Morris at outcome (SD)	4.6 (4.0)	6.8 (4.5)	9.5 (6.3)	5.4 (4.8)
Modified Oswestry Score (SD)	18.3 (11.1)	25.4 (12.0)	29.2 (21.0)	16.7 (10.9)

Table 7.6: Baseline characteristics and outcome data (Dascanio et al. 2014)

A group of musculoskeletal physiotherapists with appropriate additional training delivered the trial interventions. Ten weekly treatment sessions were delivered, treatment sessions of individual interventions were 30 minutes in duration.

The Roland Morris Disability Questionnaire (RMDQ) and the Modified Oswestry Disability Index (MODI) Questionnaire was used. Analysis of covariance (ANCOVA) was conducted to obtain treatment effects. A zero attrition rate was achieved with no dropouts following randomisation. The

study was not powered to detect a significant difference; however the manual therapy intervention group gained the most improvement across the time points (statistical significance was not achieved).

Kumnerddee (2009): conducted a randomised comparative study to provide preliminary data. Their subjects were 18 male Thai Military personnel aged 20 - 40 years, who had experienced myofascial back pain in the preceding three months. No information regarding the randomisation technique was provided. (9 Thai massage, 9 Acupuncture). The baseline characteristics and outcome data are presented in table 7.7.

Parameters	Thai Massage N = 8 (mean ± SD)	Acupuncture N = 9 (mean ± SD)	P-Value	95% CI
Age (years)	26.25 ± 6.84	29.00 ± 6.84	0.42	-0.43, 9.83
Male (%)	100	100	-	-
Onset (weeks)	12.78 ± 22.71	14.81 ± 22.73	0.86	-21.49, 25.55
McGill scores At baseline	16.13 ±8.94	15.78 ± 8.41	-	-
VAS (mm) At baseline	4.56 ± 1.37	4.19 ± 2.70	-	-
McGill scores At outcome	10.25 ± 11.02	2.11 ± 2.11	0.02 (TM) 0.002 (A)	1.01, 10.74 6.91, 20.42
VAS (mm) At outcome	2.15 ± 2.61	0.45 ± 0.71	0.03 (TM) 0.003 (A)	0.32, 4.50 -2.63, 1.88

Table 7.7: Baseline characteristics and outcome data (Kumnerddee, 2009)

The interventions included; a) Traditional Thai Massage (TTM), which is a form of deep massage said to manipulate the local and peripheral tissue. One hour of treatment was provided with a qualified therapist. b) acupuncture delivered seven set points plus additional points as the therapist felt appropriate to the patient. No information was provided on the length of treatment or needle type or point location. A single acupuncturist was used to deliver the intervention. Five sessions were delivered every two – three days over a ten-day period.

Blind measurements were collected at baseline, day three, day eight and at the end of treatment. The Thai version of the short form McGill pain questionnaire and a visual analogue scale (VAS) were used, descriptive statistics were provided, and a paired T-test was used. One participant dropped out of the TTM group of the study, no information regarding ITT analysis was given. Results reported a statistically significant improvement post intervention in the acupuncture group. A trend of improvement for both interventions and at all time points was noted but not statistically significant.

Giles et al. (2003): conducted an RCT comparing medication, acupuncture and spinal manipulation for individuals over the age of 17 years, suffering spinal pain for more than 13 weeks. 115 spinal participants were recruited from a hospital multidisciplinary spinal unit. Randomisation occurred via 150 well-shuffled envelopes.

Two acupuncturists and one chiropractor provided the interventions, which were delivered for 20 minutes twice weekly, for a maximum of nine weeks. For acupuncture 8-10 needles were used, manipulation involved high thrust techniques and medication included non-steroidal anti-inflammatory medication.

The Oswestry scale was used as the main outcome measure. The results were separated into cervical and lumbar spine pain, therefore data set for the lumbar spine pain only could be used for this review. Attrition was high, out of the 115 spinal patients randomised; only 69 reached the study end point. Additionally 36 participants changed treatment during the trial due to feeling their intervention was ineffective or due to side effects, though an ITT analysis was used. Pre and post intervention changes were analysed using the paired Wilcoxon signed-rank test and comparisons between the groups were analysed using the Mann-Whitney U test. The manipulation group gained greater improvement for lumbar pain participants. The baseline characteristics and outcome data are presented in table 7.8.

Parameters	Manipulation N = 35	Acupuncture N = 34	Medication N = 40	P-Value
Age (years)	38	35.1	36.7	0.66
Male (%)	51.4	55.9	57.5	0.86
Onset (years)	8	6.5	7.3	0.77
Oswestry Disability Index – median at baseline	24.0 (10 - 36)	27 (11 - 36)	32 (23 – 49)	0.06
VAS (mm) Median at baseline	5.0 (4 - 8)	6 (4 – 8)	32 (3 – 8)	-
Oswestry Disability Index – median at outcome	12 (0 – 29) P = 0.01	24 (11 - 36) P = 0.02	0.0 (11 – 46) P = 0.22	-
VAS (mm) Median at outcome	-2.5 CI -5, -21 P = 0.0001	4 CI -1, +2 P = 0.33	5 (2 – 7) P = 0.77	-

Table 7.8: Baseline characteristics and outcome data (Giles et al. 2003)

Cherkin et al. (2001): conducted a randomised trial of 262 participants, comparing the effectiveness of acupuncture, therapeutic massage and self-care education for back pain. Participants were invited to participate six weeks after a consultation in primary care for LBP. 693 consent forms were returned, but only the first 262 20 - 70 year olds were enrolled. Randomisation was conducted via computer assisted telephone interview (94 Acupuncture, 78 Massage, 90 Self-care). The baseline characteristics and outcome data are presented in table 7.9.

Parameters	Massage N = 78	Acupuncture N = 94	Self-care N = 90	P-Value
Age (years) (SD)	45.7 (11.4)	45.3 (11.5)	43.8 (11.7)	-
Male (%)	31	48	44	-
Onset (years)	-	-	-	-
Roland Morris at baseline (SD)	11.8 (4.4)	12.8 (5.3)	12.0 (5.3)	-
Roland Morris at outcome (95% CI)	6.3 (5.1 – 8.1)	7.9 6.5 – 9.3	8.8 7.4 – 10.2	0.001 (adjusted)

Table 7.9: Baseline characteristics and outcome data (Cherkin et al. 2001)

12 massage therapists and seven acupuncturists were used for the study, massage followed a protocol of soft tissue massage for the massage group and basic TCM needling techniques with electrical stimulation, cupping, heat and exercise prescription were permitted for the acupuncture group. The self-care group received a book and two professional videos of self-care of LBP.

The Roland Disability Scale was used as a primary functional measure. An intention to treat analysis was used, using analysis of covariance (ANCOVA) 80% power to detect a 2.5 difference. The massage intervention was favoured after ten-weeks, statistical significance was achieved comparing massage to acupuncture and self-care.

Giles et al. (1999): piloted a three-armed study of spinal pain, comparing acupuncture, a nonsteroidal anti-inflammatory drug and spinal manipulation. They recruited 77 participants from a hospital spinal unit; the participants were required to be of 18 years of age and suffered spinal pain for at least 13 weeks. Participants were randomised to one of the three groups, randomisation occurred via 150 well-shuffled envelopes (32 manipulation, 16 Acupuncture, 20 Medication). The baseline characteristics and outcome data are presented in table 7.10.

Parameters	Manipulation N = 32	Acupuncture N = 18	Medication N = 19	P-Value
Age (years)	42	44	3	0.19
Male (%)	53	35	37.5	0.38
Onset (years)	8	6.5	7.3	0.77
Oswestry Disability Index – median at baseline	28.0	24	20	-
VAS (mm) Median at baseline	5.0	4.3	3.5	-
Oswestry Disability Index – median at outcome	-8.5 CI: -14, -4 P = 0.0004	+0.5 CI: -8, +11.8 P = 0.77	0.0 CI: -4, 0 P = 0.71	Note: change seen = (before minus after intervention)
VAS (mm) Median at outcome	-2.5 CI -5, -21 P = 0.0001	+0.8 CI -1, +2 P = 0.33	+0.3 CI -2, +0.2 P = 0.34	Note: change seen = (before minus after intervention)

Table 7.10: Baseline characteristics and outcome data (Giles et al. 1999)

Interventions were described and participants received either six 15 – 20 minute treatments of Manipulation over a three to four week period, or six 20-minute treatments of acupuncture over a three to four week period. (8-10 needles were used, and electrical stimulation was applied). Or in the medication group, non-steroidal anti-inflammatory medications were prescribed. Four acupuncturists were used for acupuncture intervention and one resident chiropractor was used for manipulation intervention.

The Oswestry scale was used as a primary functional measure, and an unpaired Kruskal-Wallis test was used. The results were separated into cervical and lumbar spine pain, therefore data for the lumbar spine only could be used for this review. Dropout rate was high in this study and varied across groups (26% Manipulation, 50% Acupuncture, 20% Medication) but all participants were followed up by telephone. The spinal manipulation group only achieved a statistically significant improvement.

7.12 Adverse events

Table 7.11 summarises the adverse events reported across the included studies. Only minor adverse events were reported, which encompassed post treatment soreness (Kumnerddee (2009); Cherkin et al. 2001), with no major adverse events reported. Some minor side effects were also reported in the medication arms of two trials (Giles et al. 1999; Giles et al. 2003).

Study	Sample size	Methods used to assess adverse events	Adverse events assessed	Reported adverse events
Dascanio et al. (2014)	59	Not reported	All adverse events	No adverse events reported
Kumnerddee (2009)	18	Self-reported	All adverse events	One minor adverse effect considered not serious: One participant dropped out of the massage group due to post massage soreness
Giles et al. (2003)	115	Self-reported	All adverse events	No serious adverse events reported No minor adverse effects reported in acupuncture or manipulation arms. Seven minor side effects reported in the medication arm
Cherkin et al. (2001)	262	Self-reported	All adverse events	No serious adverse effects reported, some minor effects occurred: 11% acupuncture group and 13% massage participants reported significant discomfort during or shortly post intervention
Giles et al. (1999)	77	Self-reported	All adverse events	No serious adverse events reported No minor adverse effects reported in acupuncture or manipulation arms. Three minor side effects reported in the medication arm

Table 7.11: Adverse events reported across the included studies

It was unclear across all the studies if registration of the adverse events was systematic. Uncertainty exists therefore on the reliability of the reporting of adverse incidents.

7.12.1 Methodological quality including the risk of bias in the included studies

The results of the methodological quality and risk of bias assessment are presented in table 7.12. The CBRG 12 criteria forms (appendix E14) were used to calculate the quality and risk of bias (Furlan, 2015).

Study	No of Subjects	Randomisation method	Bias assessment score	Risk of Bias	Quality
Dascanio et al. (2014)	59	Computer generated	A + B + 11/12	Low risk	High
Kumnerd dee (2009)	18	Randomised (no information reported)	6/12	High risk	Low
Giles et al. (2003)	115	Shuffled envelopes	A + B + 9/12	Low risk	Medium
Cherkin et al. (2001)	262	Computer generated	A + B + 9/12	Low risk	High
Giles et al. (1999)	77	Shuffled envelopes	A + B + 6/12	High risk	Medium

Table 7.12: Results of methodological quality and risk of bias assessment

Table 7.12 shows Kumnderddee's (2009) study was of high risk of bias and low quality. It was a very small-scale study of men only, with a poor methodological quality. No information was provided for the randomisation and allocation concealment process, it considered pain of

three months or less duration (acute – sub-acute), its primary outcome assessment occurred at 10 days following the first of five intervention treatments, no control arm was used, it used pain scales for its outcome measures and its subject numbers did not achieve the minimum of 12 required for a pilot study (Julious, 2005). Due to the poor methodological quality including the differing time scales, the outcomes measurements used, and the high risk of bias, this study was excluded from the planned meta-analysis pooling of data.

Table 7.12 shows one study was high risk of bias and medium quality (Giles et al. 1999). One study was of low risk of bias and medium quality (Giles et al. 2003). Two studies were low risk of bias with high quality (Cherkin et al. 2001; Dascanio et al. 2014).

All five studies were randomised. One study provided no information of the randomisation process (Kumnderddee, 2009). Envelope concealment randomisation was used in one of the studies and one pilot (Giles et al. 2003; Giles et al. 1999). Two studies used computer-conducted randomisation (Dascanio et al. 2014; Cherkin et al. 2001).

None of the trials were blinded due to the pragmatic nature of the trials. Blinding of the outcome was reported in all five of the studies.

Giles et al. (1999) and (2003) also reported very high attrition and crossover rates, which was not observed in the other studies, however ITT analysis was used. One pilot study presented an attrition rate of zero post randomisation, (Dascanio et al. 2014).

Four studies were assessed to be of medium and high quality (Dascanio et al. 2014; Giles et al. 2003; Cherkin et al. 2001; Giles et al. 1999), all using a functional outcome measure and therefore these studies were included in the meta-analysis.

7.13 Adequacy of interventions

In all studies selected for this review the interventions were considered adequate for the population they targeted. Table 7.13 shows the interventions provided for each of the selected studies.

It was the view of both reviewers that the interventions across all the studies were delivered by appropriately qualified practitioners (Chartered physiotherapist, acupuncturist, chiropractor, massage therapist) and provided effective intervention types (e.g. manual therapy, massage, manipulation, acupuncture), with suitable numbers of treatment (e.g. 10, 12, 6 & 5), appropriate length of treatment sessions (30-45minutes, one hour, 20minutes) and appropriate treatment techniques delivered.

Regarding the time period the treatment was delivered for however, one study did not fit within the time frame of the others. Kumnerddee's (2009) intervention was delivered over a ten-day period, which was considered a very short intervention period by both reviewers and the third reviewer. It was considered that minimal outcomes would be understood over such a short treatment parameter, which did not reflect current clinical practice. The other four studies (Dascanio et al. 2014; Giles et al. 2003; Cherkin et al. 2001; Giles et al. 1999) treatment period ranged from 4 weeks to 12 weeks, which was considered more in line with standard clinical practice.

Study	No of Subjects	Interventions	Intervention details	Adequacy
Dascanio et al. (2014)	59	Acupuncture Manual Therapy Combined A and MT Usual care	Ten weekly sessions over a maximum of twelve weeks. Delivered by experienced physiotherapists Western acupuncture was delivered incorporating some traditional principles 30 minute sessions Manual therapy 30 minute sessions Combined A & MT received 45 minute sessions No details provided	High adequacy
Kumnerddee (2009)	18	Thai Traditional Massage Acupuncture	Five sessions delivered over a ten-day period. One-hour massage was delivered, Chinese acupuncture for myofascial pain. No information on treatment time for acupuncture. 7 needles used as protocol plus additional needles no information provided.	Moderate adequacy
Giles et al. (2003)	115	Manipulation Acupuncture Medication	Maximum of nine weeks of treatment Manipulation occurred twice weekly high-velocity, low-amplitude trusts to spinal areas Acupuncture sessions occurred twice weekly 8-10 needles for 20 minutes plus 5 needles for five minutes Medication two visits to physician prescribed Celebrex (200-400mg, doses at the discretion of the physician)	High adequacy
Cherkin et al. (2001)	262	Acupuncture Massage Self-care	Acupuncturists and massage therapists were able to schedule up to ten sessions over ten weeks. TCM acupuncture was used, decisions of number and location of needles was left to the provider, thought he acupuncturists established a treatment protocol with the trial consultants. A protocol of massage was developed by the masseurs and consultants, acupressure and shiatsu were not permitted but all other styles of massage were permitted. The self-care group received high quality information including a back book and two self-management videotapes.	High adequacy
Giles et al. (1999)	77	Manipulation Acupuncture Medication	Manipulation = 6 treatments over 3 -4 weeks, 15 – 20 minute appointments Acupuncture = 6 treatments over 3 -4 weeks, 20 minute appointments, 8-10 needles were used, electrical stimulation was also applied Medication = non-steroidal anti-inflammatory, no further details provided	Moderate adequacy

Table 7.13: Adequacy of interventions

7.13.1 Clinical relevance

The interventions selected within this review are commonly used interventions for LBP within clinical practice; therefore determining their clinical relevance, effectiveness and indication for use would be very useful. Table 7.14 highlights the findings of clinical relevance assessment for each study.

Item	Dascanio et al. (2014)	Kumnerddee (2009)	Giles et al. (2003)	Cherkin et al. (2001)	Giles et al. (1999)
Based on the data provided, can you determine if the results will be clinically relevant?	Yes	Yes	Yes	Yes	Yes
Are the patients described in detail so that you can decide whether they are comparable to those that you see in your practice?	Yes	Yes	Yes	Yes	Yes
Are the interventions and treatment settings described well enough so that you can provide the same for your patients?	Yes	Yes	Yes	Yes	Yes
Were all clinically relevant outcomes measured and reported?	Yes	No	Yes	Yes	Yes
Is the size of the effect clinically important?	No	No	Yes	Yes	No
Are the likely treatment benefits worth the potential harms?	Yes	Yes	Yes	Yes	Yes

Table 7.14: Clinical relevance assessment

All five studies provided data, participant information and appropriate details of the intervention parameters and settings to allow the interpretation of relevance to practice. Kumnerddee's (2009) study only reported pain scales, which have been shown to be subjective and less reliable as an outcome tool, comparative to functional measures outcomes

(British Pain Society, 2019), the other four studies reported using functional outcome measures.

The effect size was not considered clinically important in Kumnerddee (2009) study and the two pilot studies (Dascanio et al. 2014; Giles et al. 1999) primarily due to lack of power to detect a difference within the studies (Schober et al. 2018).

All five studies reported treatment benefit (though not statistically significant for the three smaller studies (Dascanio et al. 2014; Kumnerddee, 2009; Giles et al. 1999) and as presented in section 7.12 none of the studies reported any serious adverse events.

The evidence from this systematic review was limited primarily due to the small study sizes, therefore it is not possible to provide distinct recommendations of clinical relevance. However the meta-analysis results of the included studies indicated manual therapy showed an advantage over acupuncture.

7.14 Meta-analysis

Four of the five RCTs were selected for the meta-analysis (tables 7.15, 7.16, 7.17, 1.18). The data required conversion into means and standard deviations through statistical calculation for the Giles et al. 1999 and Giles et al. 2003 studies (appendix E16). Different outcome measures (RMDQ and ODI) were used, therefore pooled estimates of effect were provided as standardised mean differences for the analysis.

Outcome measure	Intervention	Mean (SD)	Sample size
RMDQ	Acupuncture	7.9 (6.86)	94
RMDQ	Massage	6.3 (7.36)	78

Table 7.15: Mean (SD) scores for Cherkin et al. (2001)

Outcome measure	Intervention	Mean (SD)	Sample size
RMDQ	Acupuncture	6.8 (4.5)	14
	Manual therapy	4.6 (4.0)	16
MODI	Acupuncture	25.4 (12.0)	14
	Manual therapy	18.3 (11.1)	16

Table 7.16: Mean (SD) scores for Dascanio et al. (2014)

Outcome measure	Intervention	Mean (SD)	Sample size
MODI	Acupuncture	24 (8.21)	18
MODI	Manual therapy	28 (20.77)	32

Table 7.17: Mean (SD) scores for Giles et al. (1999)

(Interquartile range converted into SD's, using the normal distribution model (Bland, 2000) (Appendix E16).

Outcome measure	Intervention	Mean (SD)	Sample size
MODI	Acupuncture	24 (19.40)	34
MODI	Manipulation	12 (17.91)	35

Table 7.18: Mean (SD) scores; Giles et al. (2003)

(Interquartile range converted into SD's, using the normal distribution model (Bland, 2000) (Appendix E16).

7.15 Meta-analysis results

The meta-analysis combined the results of four of the included studies comparing acupuncture with manual therapy, using continuous outcomes. One study was not included (Kumnerddee, 2009) due to major flaws in the study.

All meta-analyses are presented as summary effect estimates for acupuncture versus manual therapy, with standardised mean difference and 95% confidence intervals. Inverse Variance (IV) weighting was used across all the meta-analyses, to give each study a weighting determined by

how informative each study was (affected by the size of each study and the standard deviation of the outcome).

The results of the first meta-analysis are plotted in figure 7.2. I^2 was calculated at 45% thus a fixed effects meta-analysis model was used.

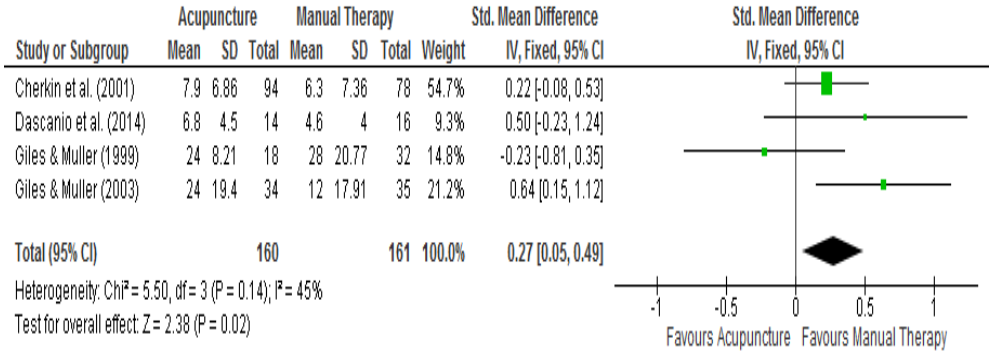

Figure 7.2: Forest plot of comparison: Acupuncture versus Manual Therapy for LBP (Outcome measures; RMDQ – Dascanio & Cherkin; ODI – Giles & Giles)

Figure 7.2 shows the fixed effect estimate for a between group comparison of acupuncture versus manual therapy, with a favourable effect towards manual therapy (0.27, 95% CI 0.05, 0.49), (P = 0.02) for this outcome.

Giles et al. (1999) was an outlier comparative to the other studies within the analysis, it is a pilot study of medium quality with a high risk of bias therefore a sensitivity analysis was conducted without this study (figure 7.3).

Figure 7.3 shows the results of the fixed effect estimate for a between group comparison of acupuncture versus manual therapy meta-analysis without Giles et al. (1999). The confidence interval of the pooled estimate (the diamond) is reduced in the forest plot and the overall combined effect size was 0.36 (95% CI 0.12, 0.60) (P = 0.004), in favour of manual therapy. I^2 was calculated at 8% thus a fixed effects meta-analysis model was used.

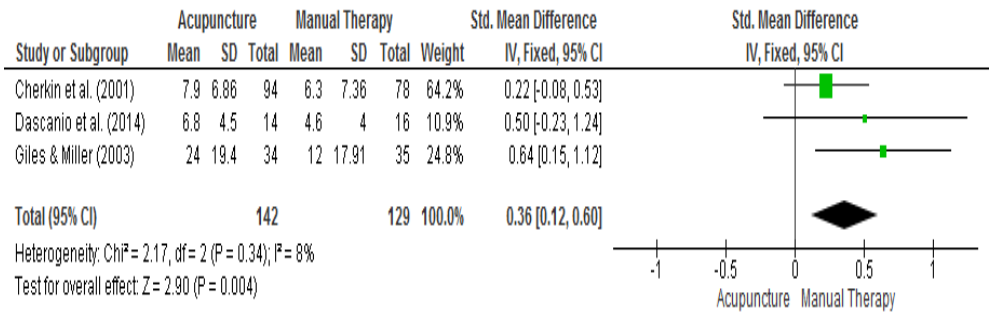

Figure 7.3: Forest plot of comparison: Acupuncture versus Manual Therapy for LBP (Sensitivity analysis minus Giles et al. 1999) (Outcome measures RMDQ Dascanio. Cherkin. ODI Giles et al. 2003)

In the previous analyses (figure 7.2 and 7.3) both Dascanio and Cherkin used the RMDQ as their primary outcome measure and Giles et al. (2003 & 1999) used the Oswestry disability index. Dascanio et al. (2014) used both the RMDQ and the MODI in their study, therefore a sensitivity analysis was able to be conducted using the MODI for Dascanio et al. (2014), thus three Oswestry measures and one Roland-Morris, to assess if any differences were seen with the use of a different outcome measure (figure 7.4 and 7.5).

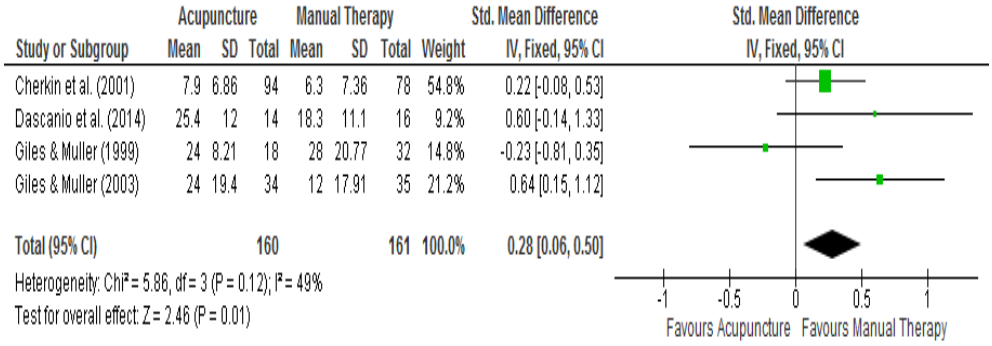

Figure 7.4: Forest plot of comparison: Acupuncture versus Manual Therapy for LBP (Outcome measure; RMDI – Cherkin; ODI – Dascanio & Giles & Giles)

Figure 7.4 shows the fixed effect estimate for between group comparison of acupuncture versus manual therapy, with a significant overall effect 0.28 (95% CI 0.06, 0.50) favourable to manual therapy, however the average effect size is slightly smaller and closer to zero / line of no effect

than seen in the previous forest plots. Statistical significance was achieved ($p = 0.01$) and I^2 was calculated at 49%, thus a fixed effects meta-analysis model was used.

Due to Giles et al. (1999) being an outlier in the analysis, a sensitivity analysis excluding the study was also conducted and is shown in figure 7.5.

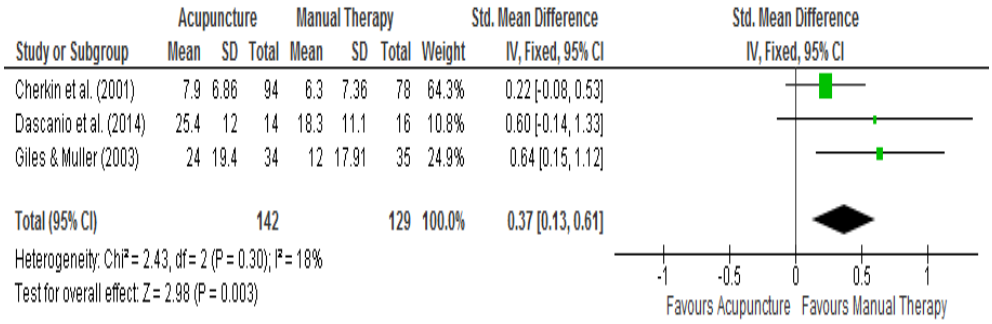

Figure 7.5: Forest plot of comparison: Acupuncture versus Manual Therapy for LBP (Sensitivity analysis minus Giles et al. (1999) (Outcome measures; RMDQ - Cherkin; ODI – Dascanio & Giles et al. 2003)

Figure 7.5 shows the fixed effect estimate for between group comparisons of acupuncture versus manual therapy, with a significant overall effect 0.37 (95% CI 0.13, 0.61) (P = 0.003) favourable to manual therapy, the average effect size is larger than seen in the previous forest plots. I^2 was calculated at 18%, thus a fixed effects meta-analysis model was used.

7.16 GRADE quality of evidence

The GRADE quality of evidence assessment was used to assess the strength of evidence. I assessed two outcomes the functional status outcome measure and adverse events. The evidence profile, the certainty assessment, summary of findings, and relative importance for each outcome are detailed in table 7.19. The PICOS criteria (section 7.4.1) were used for the identified research and the effect sizes were established from the meta-analysis (figure 7.2).

7.16.1 GRADE certainty assessment

The GRADE analysis (table 7.19) for this SR included four randomised controlled studies in the assessment (Cherkin et al. 2001; Dascanio et al. 2014; Giles et al. 2003; Giles et al. 1999).

Risk of bias: For risk of bias the rating was downgraded to 'serious' for both outcomes (table 7.19) due to the limitations of the included studies, the small sample size (due to the reduced precision of any estimated treatment effect) and two of the studies being pilot studies (Zhang et al. (19) 2018). The GRADE criteria gives a combined summary assessment for the risk of bias and thus a different risk of bias rating to that reported for the CBRG criteria in table 7.10.

Inconsistency: For inconsistency the rating was not downgraded for either outcome (table 7.19) and considered not serious, as there was relative consistency in the reported treatment effects across the studies and heterogeneity was considered low to moderate (I^2 45% figure 7.2) (Zhang et al. (20) 2018; Guyatt et al. 2011b).

Indirectness: For indirectness the rating was not downgraded for either outcome (table 7.19) and considered as not serious, as all the included studies: included a population of LBP, directly compared the relevant interventions and reported their outcomes (Zhang et al. (19) 2018; Guyatt et al. 2011c).

Imprecision: For imprecision the rating was not downgraded for either outcome (table 7.19) and considered as not serious. Certainty is reduced and a rating down of the quality occurs, if clinical choice would differ if the true effect were at the upper versus the lower end of the confidence interval (Zhang et al. (20) 2018; Guyatt et al. 2011d), which was not considered to be the case.

Other considerations: No other considerations or concerns were raised for the certainty assessment for the functional status outcome (table 7.19). However for the adverse events outcome the rating was downgraded to

'serious' (table 7.19), as there was a lack of transparency and information regarding the systematic reporting of adverse events across the included studies (Zhang et al. (20) 2018).

7.16.2 GRADE summary of findings

321 participants were included across the four studies. The overall effect size derived from the meta-analysis was 0.27 (95% CI 0.05, 0.49) (table 7.19).

The quality level of the evidence, for this review, was determined by the GRADE criteria. The quality was rated as moderate for the functional status outcome (table 7.19), this was due to one of the criteria form the certainty assessment (risk of bias; explanations given as footnotes) being downgraded to serious, therefore following GRADE's criteria the quality rating was moved down one level from high to moderate (GRADE, 2013; Schünemann et al. 2013).

For the adverse events outcome it was rated as low quality, this was primarily due to two ratings being downgraded to serious on the certainty assessment (risk of bias and other considerations; explanations given as footnotes, table 7.19), therefore following GRADE's criteria the quality rating was moved down two levels from high to low (GRADE, 2013; Schünemann et al. 2013).

7.16.3 GRADE relative importance

Consideration was given to the importance of the health benefits and potential harms of the therapies prior to making the decision and a rating of four out of nine was given on the GRADE rating scale (table 7.2), thus it was considered important but not critical for decision making (GRADE, 2013; Schünemann et al. 2013).

7.16.4 GRADE strength of recommendation

Based on the quality of evidence assessed with the GRADE strength of recommendation framework (Andrews et al. 2013), a weak recommendation can be given in favour of manual therapy over acupuncture. This recommendation was due to uncertainty remaining regarding the benefits outweighing the harms and the quality assessment rated as moderate and low (Andrews et al. 2013). Limitations in the primary research and the analysis prevent a strong recommendation from being given, and it is not possible to say from this review that all persons would choose manual therapy based on the current evidence base (Andrews et al. 2013; Cochrane, 2013; GRADE, 2013; Schünemann et al. 2013).

Outcome	Certainty Assessment							Summary of Findings			Relative Importance
	No. of Studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other Considerations	No. of Patients	Effect	Quality	
Functional status **Acupuncture Vs Manual Therapy for LBP**	4	Randomised Control Trials	Serious[1]	Not Serious	Not Serious	Not Serious	None	321	0.27 (0.05, 0.49)	⊕⊕⊕◯ Moderate [1]	4/9 (Important but not critical)
Adverse Events **Acupuncture Vs Manual Therapy for LBP**	4	Randomised Control Trials	Serious[1]	Not Serious	Not Serious	Not Serious	Serious [2]	321	Nil adverse events	⊕⊕◯◯ Low [1,2]	4/9 (Important but not critical)

Table 7.19: GRADE evidence profile and summary of findings: acupuncture versus manual therapy for LBP

[1] Limitations in the studies design exists due to small sample sizes and two of the four studies being pilot investigations [2] Lack of transparency and information regarding systematic reporting of events

7.17 Discussion

Five studies engaging 531 participants were included in this systematic review. The methodological quality of the selected studies was high for three studies and low for two studies, however none of the studies are large scale recent RCTs. The selected studies were heterogeneous in their selection of study population, manual therapy and acupuncture intervention administered, number of intervention arms and controls, outcome measures, follow up time points and the presentation of the results.

While manual therapy is considered a global term for a variety of hands on treatment for LBP (NICE, 2009) in practice this led to heterogeneity within this review, with three differing types of manual therapy delivered (massage, manual therapy, manipulation), however for the purpose of this systematic review they were considered similar enough to combine the results and for a meta-analysis.

In addition the global term acupuncture is considered as 'the insertion of needles' (WHO, 2002) and thus a variety of acupuncture intervention techniques were delivered across the studies, though heterogeneity existed, for the purpose of this systematic review they were considered similar enough to combine the results and for a meta-analysis.

If a greater number of research studies had existed in this area of study, sensitivity analyses could have been used to account for variation in manual therapy and acupuncture techniques and thus inform the results further.

The time frame of the condition was another area for consideration for this review. Four of the included studies selected participants with sub-acute and chronic low back pain; and within this patient group there appears to be some evidence to suggest manual therapy is superior to acupuncture across these studies, and this was supported by the meta-analysis results.

One of the smaller trials considered pain of three months or less therefore considering acute and sub-acute LBP (Kumnerddee, 2009), however as it was a preliminary data trial very small sample size of 18 participants were included in the study. No control arm was included in this study and the analysis was conducted at ten days, therefore regression to the mean may have played a role in any improvement seen in the participants within this study (Torgerson and Torgerson, 2008). While acute LBP is included in various guidelines it is also reported that acute and sub-acute LBP often resolves spontaneously for the majority of individuals (Qaseem et al. 2017). While this author was contacted for further information or a follow-up study, no response was received at the time of writing. Minimal conclusions could be drawn for acupuncture versus manual therapy for LBP in its acute phase from the Kumnerddee (2009) study.

The meta-analysis and forest plots all presented a favourable effect towards manual therapy, with consistent average effect sizes (0.27, (95% CI 0.05, 0.49), 0.36 (95% CI 0.12, 0.60), 0.28, (95% CI 0.06, 0.50), 0.37, (95% CI 0.13, 0.61)). The meta-analysis results are consistent with the findings of the four individual studies.

Caution is required in interpreting the results of the meta-analysis as though the studies were of good or medium quality, they were small studies with two being pilot investigations (one a high risk of bias (Giles et al. 1999) and three of the studies being over 15 years old (Giles et al. 2003; Cherkin et al. 2001; Giles et al. 1999).

The GRADE assessment offered only a weak recommendation in favour of manual therapy, due to limitations and uncertainty in the primary research, and due to a low to moderate quality assessment. Further high quality, large full-scale studies and analysis are required to draw firm conclusions in this area of research.

Limitations:

A number of limitations exist within this systematic review. Principally the review protocol was not registered or published prospectively, this

occurred primarily due to limitations and time scales of this PhD. While considerable effort was made to ensure the integrity of the review, lack of publication leaves it subject of criticism due to lack of transparency, and the questionable opportunity to deviate from the original protocol exists. The opportunity for conscious or subconscious influences on the review from the data extracted could not be refuted (PROSPERO, 2019), limiting this review comparative to the rigorous standards expected of systematic reviews. Registration of protocols reduces the potential for bias, including outcome and reporting bias. It allows transparency of the planned review including a public record of the inclusion criteria and intended outcomes, and in addition it also avoids duplication of systematic reviews (Misra and Agarwal, 2018; Stewart, Moher and Shekelle, 2012).

Additionally this systematic review planned only to consider primary outcomes using pain and function for selection for the review, it did not plan to collect data on economic outcomes, patient satisfaction, adverse reactions, negative consequences of the interventions, side effects, recurrence, fear avoidance behaviours, medication, depression. This limited the potential knowledge gained from this systematic review and limited the author's ability to make informed decisions regarding the included studies.

In order to complete the review effectively information regarding adverse events was required for the clinical relevance analysis and GRADE's summary of findings analysis. Thus the authors deviated from the original planned protocol to collect data on adverse events, allowing informed decisions on whether the treatment benefits were worth any treatment harms, (table 7.11) and to grade the level of evidence (table 7.19). Deviation from the original protocol is not ideal, however detail of why and how this was conducted is provided to ensure transparency.

The protocol inclusion criteria specified an age range of 18 to 65 years, however in a deviation from the protocol it was decided if a publication included individuals from the age of 16 years or over the age of 65 years who were not suffering comorbidities then these studies would not be

excluded based on their age specification but considered for inclusion. This decision was made based on a change in the age demographics in the update of the NICE LBP guidelines (2016). No other deviations from the planned protocol were undertaken.

The limited number of RCTs available with a low risk of bias in this area is a limitation of this study, as is the ambiguity surrounding the impact of publication bias, both noteworthy limitations but those which are perhaps inherent to many studies. Significant effort was made to find all published and unpublished RCTs, however some studies may have been missed due to the databases searched being predominantly UK based and the English language only. Further consideration would need to be given to not having a language restriction for future systematic reviews in this area. Additionally the restriction of search strategy strings inputted for this systematic review may not have been comprehensive enough and limited the studies isolated, thus limiting the review.

The heterogeneity across the included studies was another limitation of this review; while every careful consideration was given to the inclusion criteria of each study it was not an issue that could be resolved for this systematic review. The data collected from the included studies was also limited to a twelve-week outcome measurement post intervention completion, therefore limitations to the report of long-term effects of acupuncture or manual therapy for patients with LBP are apparent, long-term follow-up of studies should be considered for future reviews in this area. If any future studies were discovered this review would be updated.

7.18 Implications for clinical practice and future research

There is some evidence from this systematic review and meta-analysis to indicate manual therapy is superior to acupuncture, however with only a weak recommendation in favour of manual therapy from the GRADE. The conclusions are constrained by the limitations of the primary research studies available for the review. Minimal research was uncovered through

this systematic review; with only five studies investigating acupuncture versus manual therapy found and two of those were pilot studies and one a very small-scale preliminary study with limited power to establish effectiveness. Only three of the five studies were of high methodological quality and one of these studies was a pilot. Participant numbers were low across the studies and attrition across three studies was high (Giles et al. 1999; Giles et al. 2003; Kumnerddee, 2009), heterogeneity was significant with varying techniques used across the studies and three out of the four studies were more than 15 years old.

It is, therefore, not possible to reach a convincing conclusion to inform clinical practice and healthcare commissioning in this area of study, without conducting further high quality research to provide a more definitive insight.

Further investigation into the effectiveness of the various types of manual therapy available would be useful. Major differences existed in this systematic review between various types of manual therapy used from therapeutic massage of tissues through to manipulation of joints; further specification of the techniques would be a useful area for future study to establish the appropriate levels of effectiveness for each type of technique for research and clinical practice.

Additionally investigation into the range of acupuncture intervention styles would be useful, across this systematic review the intervention ranged from western medical acupuncture, traditional Chinese medical acupuncture (TCM), and Chinese acupuncture. The number and depth of needling, frequency and duration of sessions varied greatly, with some studies being prescriptive, others allowing the acupuncture provider to decide on needle location and one study using half prescribed points and half provider choice. Consideration of all these areas through research would be useful to inform future research studies and clinical practice.

7.19 Conclusions

This review provides evidence for manual therapy being superior to acupuncture for LBP, however as previously discussed caution is required in interpreting the results of the review due to the limitations of the included studies within the review. Additionally this review has determined that there is a significant lack of research in this area comparing acupuncture with manual therapy.

The evidence available from this systematic review does not support revising the guidance from the current NICE guidelines for LBP (2016). It is not clear whether acupuncture is effective for LBP or if it should be added to manual therapy, and if so, whether to recommend manual therapy or acupuncture first.

Further research is therefore required to reduce the uncertainty and inform decision-making. Serious consideration should be given to further investigation into appropriate and timely interventions for LBP. Uncertainty surrounding the efficacy of acupuncture for LBP remains and a future study of acupuncture versus manual therapy would be useful in this area of research.

A large multi-centered trial would assist in supporting the knowledge base, to advise on comprehensive treatment programs and appropriate care pathways. A large full-scale trial would aim to assist in providing clarity for the use of acupuncture for LBP.

8 Discussion of the Thesis

8.1 Introduction

The aim of this thesis was to investigate acupuncture and manual therapy in the treatment of LBP, and consider if further investigation of their use individually or in combination as a treatment package was a viable and effective option for clinical practice. Together with this, and in respect of the issues that surround the quality of previous research in this field (SAR, 2018), I wanted to establish whether a specific RCT design would be preferable for the evaluation of acupuncture and manual therapy which would satisfy guideline developers.

In this chapter I will discuss the research and:

- *Summarise and consider the key findings of this thesis*
- *Evaluate the limitations of this thesis and the conducted study, identifying the strengths, weaknesses and challenges of their completion*
- *Highlight what this thesis has contributed to the research knowledge base*
- *Relfect upon future research potentials identified from this thesis*
- *Consider what a future trial, guided by this thesis may look like*
- *Establish conclusions of the thesis*

8.2 Summary of the thesis

Chapter one:

In chapter one I introduced LBP and its non-specific nature. I described LBP's epidemiology, its impact and economic burden, its causes and prognosis. I presented current treatments available for LBP and outlined the complexities of assessing complex interventions such as acupuncture

and manual therapy, for a complex condition like LBP. I concluded with key research questions, which I considered throughout this thesis.

It has been well documented that LBP is the principal cause of global disability, and that further research to improve understanding of LBP in a variety of settings is required to improve the knowledge base (Buchbinder et al. 2018; Hoy et al. 2014). As I identified in this chapter, the lack of diagnostic clarity of LBP may well be impeding therapeutic strategies. Patients present in clinic with substantially different LBP symptoms; however they are often grouped together for diagnostic and research purposes. The one size fits all model has thus not aided information, recovery or prevention of long term LBP.

Further understanding of the subsets or types of LBP would be useful, separated into their anatomical cause if possible. Continuing to use the broad term LBP to diagnose so many, may adversely impact upon the success and development seen in research. Further investigation should be conducted into diagnostics of LBP to aid our understanding. Greater understanding of this multi-faceted condition may then inform appropriate treatments and management. Despite extensive research into the treatment of LBP, debate remains surrounding which interventions are appropriate, and which may constitute a long-term solution to the problem (Foster et al. 2018).

Identifying which causes and risk factors lead LBP to establish as chronic condition, remains a real need and challenge for the research and clinical community (NICE, 2016). Better understanding of the risks (as the new NICE guidelines recommend) through the use of assessments such as STarT Back screening tools and stratified care (Foster et al. 2014; Hill et al. 2011) may help future understanding. Through chapter one I presented the many unknowns of LBP and where research focus is required to reduce the impact of LBP and inform clinical practice.

Chapter two:

In chapter two I conducted a literature review of systematic reviews of acupuncture and manual therapy for LBP and the research comparing or combining acupuncture and manual therapy. I also reviewed the systematic reviews of international clinical practice guidelines for LBP. The international guideline recommendations for acupuncture and manual therapy were summarised and I examined their varying recommendations. I also reflected upon how they compared to the UK NICE guidelines and the changes that have occurred over time. The NICE LBP guidelines from 2009 and 2016 were reviewed and compared, with specific analysis of the evidence base for manual therapy and acupuncture. The removal of acupuncture from the recent 2016 NICE guidelines and uncertainty surrounding its provision was discussed, and I considered if acupuncture and manual therapy were appropriate interventions for LBP.

One of the significant themes discussed within this chapter was how the same international evidence, was interpreted differently in each country which produced guidelines and their guideline committee members. Thus producing divergent national clinical practice guideline recommendations. I found differing interpretation of the evidence and how it was applied to the guidelines. The choice between efficacy and effectiveness of interventions in the assessment processes of the guidelines was evident, particularly for acupuncture recommendations in the UK.

A House of Lords science and technology report in the UK gave a criterion for the evidence needed for clinical practice, it stated that:

> *"…any discipline whose practitioners make specific claims…to treat specific conditions should have evidence … above and beyond the placebo effect"*

(House of Lords (HoL), 2000)

The evidence for acupuncture was interpreted differently by the NICE 2009 GDC and the NICE 2016 GDC. NICE (2009) GDC considered non-

placebo RCTs including research comparing acupuncture to usual care and other interventions. However, with a change of GDC personnel for NICE 2016 and a change to the accepted value of the Minimal Clinically Important Difference (MCID) to 1.0 above sham. The GDC determined that the acupuncture studies were not able to satisfy the MCID or the House of Lords criterion (2000) to produce evidence above and beyond the placebo effect, and thus NICE (2016) withdrew acupuncture from its recommendations. However, as discussed in chapter two and chapter three, it may be that acupuncture is merely an effective placebo or it may be that there is evidence and recommendation for not comparing acupuncture to sham controls in trials, but using an alternative effectiveness trial. Additionally the potential bias of individuals and their specialities on guideline development committees remains a criticism of the committees. More needs to be done to develop a framework and agreement on the interpretation of the international evidence and assessing all evidence with the same framework and expectations, for a more consistent approach to the future development of clinical practice guidelines.

I reported a significant lack of high quality research for acupuncture and manual therapy from the literature reviews and clinical practice guidelines I reviewed. I established that further high quality robust research was required to establish the effectiveness of both acupuncture and manual therapy for LBP.

Chapter three:

In chapter three I reviewed the strengths and weaknesses of RCT design methodology and its relevance to a study of acupuncture and manual therapy.

Following a critique of the quality of past literature (many low and medium quality research studies) for acupuncture and manual therapy, I discussed the type of trial design, which could be most appropriate for a complex condition (LBP) with complex interventions (acupuncture and manual therapy). I considered various types of trial design, including a discussion

on previous placebo trial designs of acupuncture and manual therapy used to inform the NICE (2016) guidelines, to establish which design could be robust and potentially reduce the biases seen within the previous trials.

I presented the rationale for the selected use of a cohort design with a nested factorial RCT for acupuncture and manual therapy. Exploring its use for recruitment and retention in trials and giving justification for the use of an active control group rather than a placebo arm, thus providing a robust design with the potential for a high quality methodological study. While this design did not satisfy the House of Lords criterion (2000) (due to not including a placebo comparator, as detailed above), until a true inert placebo intervention can be developed for acupuncture, comparing interventions to one another, in combination and to usual care, within one RCT may provide a robust solution.

Having an appropriate comparator as an alternative in effectiveness trials equally requires consideration. If a comparator arm is less favoured and leads to resentful demoralisation then an imbalance of attrition could occur, and potentially lead to an inaccurate reporting of the treatment effect within those trials if not analysed appropriately. Exercise for example has questionable use as a comparator for evaluating the effectiveness of interventions for LBP, as a trend of indifference, reduced compliance, and higher attrition in the exercise arm has been demonstrated. Exercise by its nature takes additional effort for a participant compared to a treatment delivered within a clinic, and thus needs careful consideration before being used as a comparator for future studies of acupuncture or other complex interventions.

In chapter three I highlighted the need for further robust research of acupuncture and manual therapy for LBP, and proposed the use of a cohort study with a nested factorial RCT as the most appropriate trial design for a pilot study for this thesis.

Chapter four:

In chapter four I gave justification for the development of a pilot study and detailed the aims, objectives, methods and procedures of the planned pilot cohort with nested factorial RCT. The key objectives of the pilot study were to provide knowledge to inform and to investigate the practicality of conducting a future full-scale trial.

Explanations for the selected objective measures and the use of only chartered physiotherapists in the pilot study were also given. The recruitment rate targets, study documentation, study population, and inclusion and exclusion criteria with justifications were presented. The randomisation process, trial interventions, monitoring of participants and methods of analysis were all outlined. I published the protocol for this study in a peer reviewed journal (appendix A3) and the planned pilot study was delivered.

The Limitations of the pilot study are discussed below in section 8.3.

Chapters five and six:

In chapter five I reported the results of the pilot RCT and the use of the cohort study, providing a descriptive review of the results, the participant data gained and the outcomes of the trial. In chapter six I discussed the presented results.

While many opportunities exist to reduce attrition, shifting attrition to a pre-randomisation phase would aid the reliability and internal validity of studies, as it would potentially lead to less attrition once randomisation and the study had commenced, as was seen in the pilot for this study.

Meeting the expectations of participants is important and, while the participants who moved beyond three months in the study completed all paperwork and had excellent attendance rates, further communication with participants to inform them of the required input for the study at the consent point would have been useful. A telephone call following initial consent, for example, may have encouraged and supported potential

participants to enter the study and may also have provided the opportunity to collect qualitative information as part of a process evaluation.

The conducted pilot study demonstrated:

- *Recruitment rates from GP practices to the cohort study was an effective recruitment method*
- *Recruitment rates from the cohort study to the nested factorial RCT was an effective method of recruitment to the nested trial*
- *Randomisation occurring at three months reduced post randomization attrition to zero and facilitated an attuned cohort of participants*
- *Consent and acceptance of trial interventions was 97%*
- *Acceptance of therapist conducting the interventions was 100%*
- *The design with factorial RCT was appropriate for evaluating acupuncture and manual therapy for LBP*
- *The use of the outcome measures was appropriate*
- *A suggestion of the manual therapy intervention being superior*
- *The study parameters to inform a sample size calculation and projected study parameters for a full-scale study*

It was the first time this study design had been used to evaluate the complex condition of LBP and it would appear to be an efficient and appropriate design for this musculoskeletal condition. The results indicated the study design of this pilot to be promising, and thus should be given consideration for future research studies.

Chapter seven:

In this chapter I presented a systematic review of studies comparing acupuncture with manual therapy for LBP and its results, including a meta-analysis. This systematic review differed from the literature review presented in chapter two, as it focused upon primary research of acupuncture versus manual therapy or their combination, not SRs of acupuncture and manual therapy. It was also conducted systematically

following the PRISMA guidelines and used GRADE to assess the evidence. No studies examined the combination of acupuncture and manual therapy apart from the study conducted for this thesis. The systematic review and meta-analysis were conducted following the completion of the pilot RCT for this thesis, to enable the inclusion of its results and data.

The systematic review ensured all research was considered to inform a potential full-scale trial in this area. The results of this systematic review and meta-analysis indicated manual therapy might be more effective than acupuncture, however the quality of the included studies was limited. The lack of studies in this area of research and the results of this review and meta-analysis were further suggestive of the need for a large, multi-centred, full-scale well-conducted RCT of the presented pilot study or study in this area.

8.3 Limitations of this thesis

Duration of the PhD: The length of time, from start to finish of this PhD, presented a number of challenges, which limited the research. The primary research pilot study planned in 2010 was stimulated as a result of the publication of the original 2009 NICE LBP guidelines, which recommended acupuncture and manual therapy for LBP.

I designed and conducted this pilot trial because I felt an opportunity had been missed within the 2009 guidelines to offer a combined intervention of acupuncture and manual therapy, potentially saving the NHS money and possibly increasing the effectiveness of the treatments delivered. NICE (2009) recommended offering acupuncture or manual therapy and if one of the interventions was not effective to offer the other, thus potentially costing the NHS the expense of delivering two separate treatment courses and the cost of setting up acupuncture services throughout the NHS.

However, given the extended duration of the PhD period (due to working part time as a PhD student and extensions due to maternity leave and subsequent personal and family illness), more research was generated, and the NICE LBP guidelines were updated and published in 2016 removing acupuncture from their recommendations, after my research pilot study had been completed. The thesis therefore had to be adapted to take this into account and multiple sections updated.

The new NICE LBP (2016) guidelines added further uncertainty about acupuncture for LBP due to its withdrawal from their recommendations. This reversal of NICE's recommendation was then shortly followed by the American clinical practice guidelines which, in contrast, strongly promoted acupuncture for LBP in 2017 (Qaseem et al. 2017).

While the extended duration of the PhD work was a limitation, it did in itself present an opportunity to incorporate the new NICE guideline recommendations and advice into the discussions and compare both guidelines and their journey over time. The outcome of the updated NICE guidelines reinforced the importance of further research in this area and my trial design and pilot study remain appropriate. It lays the foundation for future research, which would be of the required high quality, to be considered by NICE.

I found the interpretation of evidence and its application to guidelines differed over time, both nationally and internationally. The choice between efficacy and effectiveness of interventions was key, especially for acupuncture recommendations in the UK.

Not conducting an interview or an in depth documentary of the different guideline approaches and how different decisions were arrived at from the same body of literature, is a limitation of this thesis. It would be a significant consideration for further study in this area and would be recommended to fully inform future research.

My review of acupuncture and manual therapy systematic reviews conducted in chapter two did not use a structured tool, such as AMSTAR

or ROBIS (Gates et al. 2018), for appraising the systematic reviews. This was a limitation of the review and a structured plan incorporating the use of an appraisal tool would need to be included into future reviews.

It is common practice to conduct a systematic review in advance of undertaking an RCT and thus completing only a review of the literature prior to my trial was a limitation. Conducting a systematic literature review later in the PhD following the completion of the pilot study enabled me to include the results of my pilot study to further inform the research. The completion of my SR towards the end of my PhD also ensured my thesis was more up to date with the passage of time and thus became a strength of my thesis.

I updated all areas of the thesis to maintain its relevance and status with the updated literature and guidelines, and I needed to rise to the challenge of maintaining the story of the thesis over the period of time and my journey on this PhD.

Pilot study: This thesis reports on the assessment of a cohort design with a nested factorial RCT in its pilot form for a six-month period. The study design has therefore not been tested in a full-scale trial over an extended follow-up period, thus limiting the information on long-term follow-up that can be realised and the impact of this pilot study and thesis.

The lack of a long-term follow up limited the degree to which the results adequately informed the planned long-term study, as it could not assess attrition levels or participant commitment passed the six-month time point. A full-scale study would plan to have a two-year cohort and follow-up period. Additionally it is unknown whether the effectiveness of the treatments would diminish with time and how this might differ between treatment groups. While a longer term study was initially planned for this pilot study, delays during the early part of the PhD and lengthy ethics and recruitment processes determined this PhD could only evaluate the study to the six-month time point, limiting any longer-term conclusions being drawn.

The pilot study engaged a GP database recruitment system, which limited its scope, only capturing individuals who had attended their GP within the preceding 18 months. It could not incorporate individuals who self-managed their LBP or who sought private healthcare without attending their GP. For a full-scale trial, consideration to a wider recruitment method to ensure inclusivity would be important. It has been suggested that LBP is significantly more widespread than is currently recorded (Greenberg et al. 2005); wider recruitment to include untreated or unreported cases of LBP would aid generalisability of the final study.

The generalisability of the pilot study is also limited as only adults between the ages of 18 – 65 years old were included. It would have been useful to extend the parameters of the study, as LBP becomes more prevalent in the aging population, therefore an opportunity was missed to gain information on this population type. The 2016 NICE guidelines consider individuals above 16 years without an upper age limit. Extending the age demographic and including LBP and sciatica for a full-scale study to fall in line with the new NICE (2016) recommendations would be valuable.

Due to the location of the PhD in the city of York and limited funds, the pilot study was conducted locally. York is a predominantly white area of the country with few very deprived populations and lacks the diversity of some other parts of the UK. A full-scale study would need to incorporate a more diverse population closely representing the population of the UK. A multi-regional study would be preferable to allow wider generalisability, and this would also need to make provision for non-English language speaker/reader participants.

The recruitment of only one physiotherapy practice remained a limitation of the study delaying the time frames achievable, and with no NHS provision represented. Future research would look to incorporate both the NHS provision and additional physiotherapy clinics.

While a pilot study was always planned, if we had achieved the sample size planned for and had the multiple trial sites envisaged it would have

aided the studies strength and conclusions, by providing more precise estimates for a full-scale study and greater generalisability of the results.

Limited participant preference was afforded in this trial and following the recommendations within the NICE (2016) guidelines participant preference should be incorporated into any future trial to mirror patient choice encouraged within practice.

Due to the factorial design of the pilot study, chartered physiotherapists conducted all interventions. In clinical practice there may be differences in clinical delivery of acupuncture and manual therapy if delivered by other therapists i.e. acupuncturists, osteopaths, chiropractors or masseurs. This study fails to consider other professional input, however great consideration was afforded to the profession selected to lead in this trial.

Loss of potential participants from the cohort for the nested RCT was 30% and this limited the pool of available participants to the RCT, due to many scoring below four on the RMDQ outcome measure. Further consideration would need to be given for a full-scale trial, as to whether this minimum entry level on the RMDQ should also be set for entry to the cohort study, to limit the number of potentially ineligible trial participants. Alternatively consideration of whether a score of four or more on the RMDQ is too high as an entry level for the treatment trial when so many scored lower, is required.

This pilot study lacked full transparent criteria for the analysis to compare the two objective measures tools, limiting the information gained, this would need to be addressed for an informative assessment in full-scale study.

Limited resource use data was collected as part of this pilot study; a full-scale trial would need to include a full economic evaluation to inform commissioners and advise on the costs of intervention delivery on a national scale. Incorporating a complex intervention as a service, such as acupuncture into LBP care provision as was recommended by the 2009 NICE LBP guidelines was estimated to cost £24,366,000 if delivered by a

newly funded service with acupuncturist practitioners (NICE, 2009b). However, incorporating a service such as acupuncture into the NHS, through an already funded and established physiotherapy musculoskeletal or pain service, within the NHS could easily be delivered with minimal additional cost (training, time and acupuncture needles being required, (Dascanio, 2015b) by physiotherapists, nurse practitioners or doctors. This lower cost delivery option was not considered by the NICE Guideline Development Group (GDG) in the 2009 or 2016 LBP NICE guidelines.

> *"The GDG considered the potentially considerable cost impact for the NHS if acupuncture was recommended and this would need to be underpinned by a strong evidence base of clinical and cost-effectiveness, which the GDG did not feel had been demonstrated."*

(NICE, 2016, page 498)

Providing full cost delivery details and economic data in a full-scale study would be essential to guide future guidelines and studies.

The conducted pilot was planned to pilot the study design for a full-scale study however there were elements of this pilot, which could have fitted into the bracket of a feasibility study. It may have been wise to extend the scope of this pilot to combine the intention of the pilot with a feasibility study to have better informed a full-scale study.

Process Evaluation: The very limited process evaluation conducted for the cohort study and the nested RCT for acupuncture and manual therapy was a limitation of this study, further analysis of the processes would have been informative. The limited process evaluation in the conducted pilot study simply considered the process of GP and physiotherapy recruitment, participant recruitment to the cohort and the nested RCT, the completion rates of the questionnaire case report forms, the success of the physiotherapists to conduct the sole and combined interventions and if any adverse events were reported. No detailed analysis however was collected on the processes of participant recruitment, ease of completion of the

questionnaires or how the interventions were delivered and if any deviations from the planned methods were undertaken. All this information would have been useful to generate knowledge to inform the planning of a full-scale trial.

To have included the collection and analysis of qualitative data; for example, the analysis of the processes of running the whole study would have highlighted more in depth the strengths and weaknesses of the study to further inform a future study.

Process evaluation information regarding the fidelity of participant journey could have further aided planning of a future trial. Feedback from potential participants on their experience of the consent process, the information sheets, the pre-screening questionnaires, and overall communication from the study, would have helped further inform decisions regarding the duel consent process and if the administration commitment of the trial was realistic for participants. A telephone interview could have been introduced to gain information regarding participant's experience of engaging with the different stages of the study and how these could have been improved for a future study.

The process and experience of; participants moving from the cohort to RCT, participants attending for their allocated intervention, seeing the therapist and their interaction with the interventions, would have been useful. Additionally information from participants not contacted to take part in the RCT and remaining in the cohort only and their experience. Gaining an understanding of participant responses and feedback would have aided preparation for planning a full-scale RCT and the process evaluation of the participants journey would need to be incorporated into a full-scale study of this pilot.

Process evaluation information from this thesis was limited in determining how the interventions were implemented in clinic by the recruited therapists. It is therefore not expressly known if what was intended, was actually delivered. Analysis of the range of manual therapy techniques

used or the choice of point selection for acupuncture treatments (including depth, needle size and point justification) was not planned or collected. Specific treatments may have been perceived as effective and thus selected by the therapists. Therapists may have selected techniques and acupuncture points they felt comfortable using, and thus patterns of treatment perhaps specific to therapists may have emerged. Information on any adaptation's therapists may have been required to make to accommodate their patients was not collected, patients' weight, fitness, perception of pain, concerns or tolerance of treatment may have all effected the patient and therapists' journey. All this information would have supported a future study on the potential use or requirement for a treatment protocol or part protocol, to ensure evidence of consistency and equity in the delivery of the interventions across the groups.

Further information on the fidelity of the therapist experience of delivering the interventions, especially in the combined intervention group, would have been useful to establish if any limitations to delivering a dual intervention existed. It would have been useful to know if the therapists felt they were delivering the interventions in the combined group as effectively as they felt they were delivering them in the solo intervention groups. How the participants were divided up amongst the therapists would have been useful to know, while the participants were allocated to their intervention independently, once the referral was sent to the clinic, the administrator determined which therapist the participant saw. Further information on the individual therapists may provide answers on whether a 'therapist effect' bias may have taken place. Further understanding would have aided planning and would need to be planned into a future full-scale study of this pilot.

A 'study' of 'this study' to evaluate the conducted processes further and how it could have been improved at each stage would have been useful to plan to run alongside this study (Moore et al. 2015; MRC, 2008; Craig et al. 2008). This process evaluation information would have aided the information derived from this study, and more extensive process

evaluation should be incorporated into a full-scale study to fully inform how the design met the objectives, the outcomes of the research and further guide future research in the area.

8.4 What this PhD contributes to knowledge

The research presented in this thesis contributes to key areas of the current research knowledge base: of RCTs nested within cohorts, of piloting a novel design, of RCTs using a factorial design to compare acupuncture to manual therapy, the combination of acupuncture and manual therapy by physiotherapists, in providing information to inform a full-scale RCT, in providing information of a preferred design for future RCTs of acupuncture and in providing current information on the research base through a systematic review and meta-analysis. I highlight below the original and significant contributions to knowledge made:

- *It was the first study to use a cohort study with nested RCT for acupuncture and manual therapy for LBP, and a factorial RCT to consider the combination of acupuncture and manual therapy for LBP, and to demonstrate it was an effective study design for the recruitment and retention of attuned participants.*
- *It showed that the cohort design with nested RCT was an appropriate design for future musculoskelatal studies*
- *It provided insight into the current research base of acupuncture and manual therapy for LBP by conducting the first systematic review in this area and highlighted a lack of research on this question*
- *It established participants willingly accept acupuncture and manual therapy and their combination as an intervention for LBP*
- *It demonstrated acupuncture and manual therapy can be combined and delivered safely by chartered physiotherapists*
- *It piloted a study, determined study parameters and informed a potential full-scale trial*

The original journal articles I have published so far from this PhD are presented in appendix A.

8.5 Implications for future research

While there has been extensive investigation of LBP over the years, few trials have been able to understand and manage the fluctuating condition of LBP. Further research should focus on the examination and diagnosis of LBP, what causes and triggers LBP, how LBP differs across a variety of settings and what the predictors are for translating an acute episode of LBP into a chronic condition.

NICE LBP guidelines (2016) in their recent publication moved away from "acute, sub-acute and chronic" categories and as an alternative considered back pain as a continuum, primarily due to many acute LBP episodes exhibiting symptoms that could be predictors of chronicity and show risk of exacerbation. A new way of thinking for LBP is thus emerging, clinical practice and research requires adaptation. Clinical practice will need to invest in the assessment and support of LBP in the acute stages for those who are identified as at risk of chronicity. Research will need to be undertaken to assess if our measurements of chronicity risk are accurate and if the interventions being delivered are effective at limiting chronicity.

Real world therapy includes multiple treatments; RCTs and guidelines would be more useful if they reflected a combined approach in order to provide more effective guidance on clinical treatment pathways. In line with the research, clinical guidelines currently recommend either / or scenarios. Consideration of an RCT with an inclusive combined approach (e.g. acupuncture, manual therapy, exercise, self-care, CBT, medication) would be advantageous to reflect an integrated care package for LBP. This integrated approach is supported by the recent NICE guidelines for LBP (2016).

The American guidelines (Qaseem et al. 2017) recommendation not offering drug therapy initially as first line treatment (recommending acupuncture and other interventions); this is long way from primary care we see practiced in the UK today. A transition period is required from guidelines to clinical practice and the incorporation of any change into medical care and General Practice is required. It will be interesting to see if the change in drug therapy recommendations will be incorporated internationally or remain specific to the USA. The reliance upon pharmacology however is becoming less desirable for many individuals within society.

Assessing the effectiveness of acupuncture in a high quality study and its use in combination with other interventions would be valuable to reduce uncertainty. It is repeatedly reported that a combined intervention approach combining acupuncture with another intervention provides additional benefit and could be more effective than current practice (Liu et al. 2015; Furlan et al. 2005). The study design presented within this thesis was effective at assessing acupuncture and could improve the quality of acupuncture research, it should thus be considered so a decision upon the effectiveness of acupuncture can be adequately determined.

Further research on GP database recruitment would be useful in order to understand differences in database system software and how they impact on selection and thus the recruitment process. Investigation and understanding of the underreporting of LBP and reasons for not consulting ones GP would be useful to the research community. Additionally recruitment of these individuals to future studies would need to be a consideration.

Further investigation into the use of the cohort study is required, consideration regarding a delay period at the start of trials, or multi-levels of administrational tasks may design out any significant levels of attrition, prior to randomisation and also before substantial trial work has been conducted, thus potentially negating the need for the cohort study. Reducing attrition in trials would improve the reliability of results and help

reduce the number of participants required for studies, whether this could be more effectively conducted via a cohort nested RCT design or a delay period and multi-level administrational design needs to be evaluated.

Investigation of the cohort of participants functioning with mild LBP (three or less on the RMDQ) would be useful to that population of individuals, and examination of the course of LBP through an extended cohort would be useful for long-term understanding of LBP.

A review of effectiveness studies using exercise, as comparators would be of interest to the research base, as attrition rates in exercise interventions are suggested to be higher and thus may give an overestimation of treatment effects of the comparative intervention. Further investigation into appropriate comparators for effectiveness studies would be useful for future research.

8.6 A potential future full-scale trial informed by this thesis

The pilot study demonstrated the cohort with nested factorial RCT design, to be feasible and effective for evaluating acupuncture, manual therapy and their combination for LBP. As presented in section 5.6 to scale-up this pilot study to a full-scale study would be achievable and there is a real need for further research in this area. The provision of more conclusive answers regarding the effectiveness of acupuncture for LBP would support knowledge, and also add to the limited knowledge base of acupuncture and manual therapy for LBP. Promoting the use of the pilot trial design, may improve the quality of research in this field; there are multiple aspects learned from conducting the pilot to take into consideration prior to planning a future full-scale RCT.

Scaling up the study: Scaling up any pilot study to a full-scale trial has challenges and repeating anything on a larger scale entails greater resources, logistics and organisation. Further explanations on the scaling up of recruitment and logistics of this pilot are also detailed in section 5.6.

Within this pilot for example, only two GP practices were used for recruitment, this would need to be scaled up considerably, identifying 20 – 25 GP practices, to recruit at least 15 GP practices to allow for adequate recruitment (as discussed in section 5.5), this would incur additional workload, logistics and cost. Enlisting primary care research networks would be a useful resource for supporting the recruitment of GPs and NHS physiotherapy practices. While there are some financial incentives for GPs to partake in research studies, and the trial comes with the benefit of providing many GP patients an intervention for their LBP, other ideas for attracting additional GPs would need to be considered for a full-scale trial. UKBEAM (2004) successfully scaled up their feasibility study from two to 14 GP practices for their full-scale study, and effectively combined the use of NHS and private practice providers to deliver the trial interventions.

Recruitment of participants with LBP as discussed (section 5.6) would need to be expanded to include non-GP practice recruitment, due to the high non-reporting rates of LBP patients (Greenberg et al. 2005; Papageorgiou, 1991). This could be done through advertising, social media platforms, recommendation, information available at private clinics, local groups, sports clubs, at other medical practices and through the NHS.

The scope of participants would be a consideration for a future study and including participants from the age of 16 years with no upper age limit (once comorbidities were excluded) would bring any future study of LBP in line with the scope of the NICE LBP guidelines (2016) and thus be recommended for a future study to ensure it was as generalisable as possible.

The interventions delivered for this pilot were conducted at one private physiotherapy practice, and with them working in close proximity, this allowed communication and consistency to be easily managed. Scaling up to multiple practices in varied cities would require a larger team to manage the logistics, training, monitoring of consistency and communication, using the national Primary Research Networks would help support this. A prescriptive intervention protocol or a partial prescriptive intervention

protocol, using the TIDieR checklist (Hoffmann et al. 2014) would be advisable to implement for therapists, if the study were scaled up, this would aid a large multi-centred full-scale study to ensure consistency, quality and equity across all the intervention groups.

Additionally incorporating NHS sites into a future study would bring with it the requirements as for other private practice sites, but in addition further NHS ethical approval at all sites would be required, logistics for participants getting to and from hospital sites and parking would need to be considered, participant perception of attending an NHS centre or a private practice would need to be considered and data collected to investigate any impact of this. As discussed previously some participants may consider treatment in private practice to be more valuable and this would be useful knowledge to explore further in the process evaluation analysis of a future study. UKBEAM (2004) did not find any significant difference in the effectiveness of interventions delivered in either private practice or the NHS, however no qualitative information was collected or analysed in their study.

Additional training and resource provision would be required within the NHS to the therapists and the support staff, ensuring recruited practitioners had adequate availability and provision within their diaries to accommodate weekly follow-up appointments. This would be essential to ensuring participants received the prescribed intervention in a timely manner. Ensuring practitioners were afforded adequate time within their busy schedules to ensure they could commit fully to the study and to treat its participants would be important, due to the already overstretched caseload of many NHS therapists. Consistent communication and additional support for therapists in the NHS may be required to fully support them.

These additional requirements and costs would all need to be factored into a future plan and budget for a full-scale trial. Due to the implication of additional costs it would be appropriate for a 'value of information analysis' to be completed, to estimate if the cost of conducting the future

research was worthwhile. A 'value of information analysis' is a quantitative process, which aims to estimate the 'return on investment' of proposed research studies, giving a value to the expected knowledge gained. It is a useful tool, which can be conducted alongside an economic evaluation in a trial and can also assess cost effectiveness of projects (Wilson, 2015).

Cohort study: The use of the cohort study should be considered; while recruitment to a cohort has previously been suggested as a potential limitation of this design (section 3.47) this pilot demonstrated it to be an effective recruitment mechanism for an RCT, with attrition occurring during the early stages of the cohort prior to randomisation to the RCT.

There was an indication from the pilot that having a run-in period and/or multiple admin layers may possibly serve to reduce attrition in a way similar to the cohort model. This would potentially reduce the additional cost and work associated with the cohort study. The cohort does however serve the purpose of inducting participants into the study and having a cohort of individuals with LBP who could also be monitored and included in potential analysis.

In addition the cohort also supported the study through the reduction of resentful demoralisation, as those allocated to the usual care arm were not informed of their allocation and continued as though they were in the cohort potentially awaiting the trial.

The cohort design also served the purpose of allowing multiple recruitment time points during the study. This could be expanded upon further in a full-scale study, allowing for entry into the trial at a later time when participants' symptoms had changed. If further recruitment to the trial was required, additional participants could be recruited from the cohort at a later time point.

The cohort recruitment model has only been piloted in this area of study and it would be useful to research to see it conducted in a full-scale study.

A full-scale cohort style recruitment model for a study of LBP would support research in this area.

Entry level to cohort study: Some alterations to the process would be recommended to reduce the loss of participants through the process of the cohort due to their ineligibility.

Introducing a minimal score of two on the RMDQ to enter the cohort would reduce the number of participants with very low-level back pain and exclude those who scored zero or one. This would streamline the recruitment to the cohort and potentially improve the number of eligible participants for the nested RCT.

Entry level to RCT: Reducing the entry level to the RCT from four to three would be indicated to reduce the number of individuals being ineligible for the study because they scored below four on the RMDQ. With the minimally clinically important difference planned to be a mean change of 1.5 on RMDQ, reducing the RMDQ score to three would still allow for change to occur and be recognised.

It was not clear if multiple low scorers on the RMDQ were due to it being less sensitive for the individuals with low level LBP, or if many individuals suffer with low levels of LBP and feel their back pain substantial enough to enter a study, even though they only scored minimally on pre-screening. Process evaluation and feedback would be planned to gain this information from a future study.

It could be argued that the cohort and the RCT should have the same entry score, however LBP varies across time and those who scored two at baseline may have a change in score at follow-up. This approach would accommodate some of those participants with persistent low levels of back pain. There would also be a differentiation between the cohort and the RCT for investigation regarding those individuals.

Factorial design: From the information gained from the pilot study a factorial trial design would be desirable in a future full-scale trial for a variety of reasons. It allows for the simultaneous comparison of more than one intervention to be compared to usual care, thus costing less than two separate effectiveness trials (Sedgwick, 2014), and it allows the study of the effect of two or more interventions conducted alone or in combination (NephJC, 2017). When designing and applying for funding for a potential future full-scale study of this pilot, it will need to be considered that time has passed, and guidelines have progressed and thus research priorities may also have changed. It may be decided for a future study that to attract funding for a trial a wider scope of LBP interventions to further reflect the NICE guidelines may need to be considered. The inclusion of a medication, exercise, other interventions and / or incorporating patient preference (as recommended by NICE 2016) would be desirable, and these would all be possible through a factorial design and it would allow some flexibility to the planned full-scale study to ensure funding was achievable.

Using a factorial design allows researchers to understand the effect of two or more independent variables upon a single dependent variable such as LBP. A standard RCT usually compares a single intervention with a control to establish effectiveness, if investigating several interventions in different trials, this would be statistically inefficient, with extensive time and work spent on participant recruitment and the burden of greater costs of running separate trials (Sedgwick. 2012). The factorial design also differs and has advantages over a multiple arm trial due to its design and statistical methods.

A factorial design allows the effectiveness of several interventions to be compared to usual care and also to all arms not including that specific intervention, the factorial design is therefore more efficient and versatile than running parallel arm trials (Sedgwick, 2012). Thus it has powerful statistical processes that allow for the analysis of the compared multiple treatments while only using one control arm, and for all participants to be included in the analyses of both investigations. Using a regression analysis

model principally provides an average of two differences, weighted according to sample size. The primary analysis is a comparison of the margins of a 2 x 2 table (see table 3.1) and the effect of each intervention is adjusted for the other interventions and any covariates. The adjustments improve the accuracy of the analysis to obtain precise estimates of effect and their standard errors (Montgomery, Peters and Little, 2003).

The factorial design is considered a 'versatile experimental design' as it allows the opportunity for researchers to examine whether the combination of interventions can affect their effectiveness. No other design allows for such information to be acquired (Sedgwick, 2012).

Factorial RCTs are most frequently powered to detect the main differences and the effects of interventions in trials; as to adequately power a factorial study to detect plausible interactions requires a significantly enlarged sample size (Montgomery et al. 2003). If research did not attract extensive cost and there was no financial implication when designing trials, then conducting a full-scale cohort with a nested factorial RCT, powered to assess interaction would be ideal. However the cost of scaling up a factorial study to test for interaction in this circumstance would be four fold (Montgomery et al. 2003), for this study, 4245 participants would be required to power it, as oppose to 1064 for a study not testing for interaction (as discussed in section 5.5.2.1), a considerable expense on an already costly study. From knowledge of funding awards for trials, it would be unlikely that funding would be granted for such a large-scale study, where costs had increased by four simply to consider the interaction of interventions, of which the effect of some was still uncertain. Maintaining the use of the factorial design (but not powering to detect interaction) would additionally give the flexibility of scaling up the planned study with future studies (following correct adaptation procedures) to potentially include the power for interaction in the future, if there was indication from the full-scale study for the need to testing this area further. The results of the full-scale study could thus be added to future investigations in the area, an example of combining the results from

research studies was conducted by MacPherson et al (2014), they performed a secondary analyse of the results from 29 acupuncture trials for a variety of conditions to explore the influence of the control group on the effect size of the interventions.

The pilot study demonstrated the factorial design to be a robust design for assessing complex interventions for a complex condition, providing a pragmatic inclusive approach, which can closely map clinical practice. The factorial design has only been piloted comparing acupuncture with manual therapy for LBP and it would be useful to research to see it conducted in a full-scale study. Using the factorial design RCT nested within the cohort would be advocated for a full-scale study of this pilot.

Long-term effects: Understanding long-term effects of interventions within research, is crucial to advocating their use and justifying their cost. Ensuring a full-scale trial considered long-term effects and incorporated long-term follow-up would be important to this area of research. In Macpherson et al. (2017b) meta-analysis they attempted to provide answers regarding the length of time the effects of acupuncture persist. They concluded the effects of acupuncture did not appear to reduce importantly at twelve months for chronic pain, with their central estimate suggesting approximately 90% of the benefit of acupuncture remained compared to the control interventions. In their recommendations they advocated for additional trials to measure long-term outcomes of acupuncture to at least 12 months follow-up and ideally beyond. Therefore long-term analysis for a full-scale study would be planned to follow-up outcomes at 12, 18 and 24 months, as this would aim to contribute new knowledge and information in this area of research.

Economic evaluation: As discussed in section 8.3 (pilot study limitations), the limited resource data collected, and the lack of economic evaluation conducted for this pilot was a limitation. Therefore a full economic evaluation of a future full-scale study would be planned and implemented. Enabling economic data to be collected from the study to inform therapists,

commissioners and guideline developers of the economic value of the investigated interventions for LBP.

Process evaluation: As discussed in section 8.3, process evaluation would be required to provide additional information and answers not provided within this pilot study. The intentions of the process evaluation of the study would intend to inform researchers, policy makers and practitioners of the processes of conducting the RCT and the assessed complex interventions. Aiming to giving insight into why some interventions are effective and why some are not, it may simply be an intervention is just not effective or it may be due to lack of equipoise or implementation in the study. If there were reasons some interventions were not seen to be effective, the process evaluation would aim to assess how the interventions and their delivery could be improved.

A future process evaluation would focus upon 1) Intervention implementation - what was actually implemented in practice and why it was done as such, were any adaptations to the intervention or its delivery required or conducted, and did the actual implementation of the interventions impact the effectiveness. 2) Causal mechanisms - close scrutiny of causal mechanisms to aid the establishment of more effective interventions and the transferability of the findings to other settings and populations. 3) Contextual factors – to understand the factors external to the intervention, which may improve or hinder the intervention (Moore et al. 2015; Moore et al. 2014).

A process evaluation of the study would be fully implemented through a study of the 'study' approach (Moore et al. 2015; MRC, 2008; Craig et al. 2008) as discussed in section 8.3 (process evaluation). A future study would use the MRC process evaluation framework for complex interventions, with focus upon planning, design, conduct, reporting and appraisal of the process evaluation of the interventions (Moore et al. 2015).

Summary: It would be proposed to undertake a full-scale cohort study with a nested factorial RCT, similar to the one undertaken for this pilot, however with significant changes implemented. Expansion of recruitment and intervention sites (including the NHS) would be conducted. The age demographic would be expanded. The entry level for the cohort would be set to a minimum of two; the minimum entry into the RCT would be lowered to three. The factorial design would be used to allow for analysis of multiple interventions but would not be powered to test for interaction due to the excessive cost implication. Additional training and support would be provided and a prescriptive or part prescriptive intervention protocol would be implemented. Long-term outcomes would be measured, a full economic evaluation would be conducted, and a full process evaluation of the study would be undertaken. These changes would aim to ensure a robust and comprehensive, high quality, full-scale study of the effectiveness of acupuncture and manual therapy, and their combination for LBP was conducted.

8.7 Conclusion

A robust evidence base is required to inform the treatment provision and policy on LBP within the UK and globally. As previously discussed, LBP is a widespread, costly and significant problem for individuals, healthcare and industry (Clark and Horton, 2018; Cholangitis 2018). There is a real need for high quality research evidence to provide more extensive and conclusive answers to the condition. It is apparent that time for a new approach to LBP is upon us, with early intervention and less reliance upon medication (NICE, 2016; Qaseem et al. 2017).

There is a significant and clear need for long-term longitudinal research into LBP and its symptomology over time. Gaining a greater understanding of predictors, the number of episodes, duration of pain, severity of disability and recovery would provide insight, to provide appropriate and timely interventions (Hoy et al. 2014).

Providing a more inclusive approach to LBP, by incorporating a variety of interventions to present a more real life experience and understanding the effectiveness of treatment combinations to reflect current clinical practice would be valuable to the area of study (Salerno et al. 2002). Perhaps similar to the combinations used in the UKBEAM (2004) study, where one intervention arm delivered manipulation followed by a course of exercise. Implementing clinical practice into research and thus informing current guidelines of a more complete approach could facilitate patients and the health services more effectively and may ultimately provide more conclusive results (Salerno et al. 2002).

There is a clear need for further research into the clinical effectiveness and the clinical necessity of current treatments for LBP. The uncertainty surrounding the use of acupuncture remains and advice on its inclusion continues to vary internationally (Qaseem et al. 2017; Wong et al. 2017; NICE, 2016; Nijs et al. 2015; Koes et al. 2010).

A large full-scale well-conducted multi-centred trial is required within this area of study. It would assist to reduce the uncertainty surrounding acupuncture and also provide further insight into the effectiveness of manual therapy and/or other interventions for the long-term benefit of patients with LBP.

I have demonstrated through the journey of this PhD and thesis, the RCT conducted, its results, the results of the systematic review and meta-analysis, and through the discussions and the conclusions, that the thesis aims have been achieved and this thesis has contributed to the knowledge base for acupuncture and manual therapy for LBP.

9 *Reference List*

Adamson, J. Cockayne, S. Puffer, S. Torgerson, DJ. (2006) Review of randomised trials using the post-randomised consent (Zelen's) design. Contemporary Clinical Trials; 27 (4): 305 – 319

Advertising Standards Authority (ASA) (2010) Advertising Guidance: Non-broadcast, Publication. September https://www.asa.org.uk/resource/publications.html

Airaksinen, O. Brox, JI. Cedraschi, C. Hildebrandt, J. Klaber-Moffett, J. Kovacs, F. Mannion, AF. Reis, S. Staal, JB. Ursin, H. Zanoli, G. (2006) European guidelines for the management of chronic non-specific low back pain. European Spine Journal; 15 (2): 192 – 300

Agence Nationale d'Accreditation et d'Evaluation en Sante (2000) Guidelines department, diagnosis and management of acute low back pain (<3 months) with or without sciatica & diagnosis, management and follow-up of patients with chronic low back pain, Paris. www.anaes.fr

AGREE (2017) Appraisal of Guidelines for Research Evaluation. The AGREE Next Steps Consortium. The AGREE Research Trust; Update: December 2017. http://www.agreetrust.org

AGREE (2009) Appraisal of Guidelines for Research Evaluation. The AGREE Next Steps Consortium. The AGREE Research Trust; May 2009. http://www.agreetrust.org

Ahn, R. Woodbridge, A. Abraham, A. Saba, S. Korenstein, D. Madden, E. Boscardin, WJ. Keyhani, S. (2017) Financial ties of principal investigators and randomized controlled trial outcomes: cross sectional study. BMJ; 356. Doi: https://doi.org/10.1136/bmj.i6770

Ahn, AC. Wu, J. Badger, GJ. Hammerschlag, R. Langevin, HM. (2005) Electrical impedance along connective tissue planes associated with acupuncture meridians. BMC Complementary and Alternative Medicine;

5: 10. Doi: 10.1186/1472-6882-5-10

Ajimsha, MS. Daniel, B. Chithra, S. (2014) Effectiveness of myofascial release in the management of chronic low back pain in nursing professionals. Journal of Bodywork and Movement Therapies; 18 (2): 273 – 281. Doi: https://doi.org/10.1016/j.jbmt.2013.05.007

Alghadir, AH. Anwer, S. Iqbal, A. Iqbal, ZA. (2018) Test-retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. Journal of Pain Research; 11: 851 – 856. https://doi.org/10.2147/JPR.S158847

Altman, DG. Bland, JM. (2003) Interaction revisited: the difference between two estimates. British Medical Journal; 326: 219. Doi: 10.1136/bmj.3

Altman, DG. Whitehead, J. Palmer, MKB. Stenning, SP. Fayers, PM. Machin, D. (1995) Randomised consent designs in cancer clinical trials. European Journal of Cancer 31A; 1934-44

Ament, SMC. De Groot, JJA. Maessen, JMC. Dirksen, CD. Van der Weijden, T. Kleijnen, J. (2015) Sustainability of professionals' adherence to clinical practice guidelines in medical care: a systematic review. BMJ Open; 5: e008073. Doi: 10.1136/bmjopen-2015-008073

Anderson, GBJ. (1999) Epidemiological features of chronic low back pain. The Lancet. August; 354 (9178): 581 – 585

Anderson, GB. Lucente, T. Davis, AM. Kappler, RE. Lipton, JA. Leurgans, S. (1999) A comparison of osteopathic spinal manipulation with standard care for patients with low back pain. New England Journal of Medicine; 341 (19): 1426 – 1431. Doi: 10.1056/NEJM1999911043411903

Andrews, JC. Schünemann, HJ. Oxman, AD. Pottie, K. Meerpohl, JJ. Alonso-Coello, P. Rind, D. Montori, VM. Brito, JP. Norris, S. Elbarbary, M. Post, P. Nasser, M. Shukla, V. Jaeschke, R. Brozek, J. Djulbegovic, B.

Guyatt, G. (2013) GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. Journal of Clinical Epidemiology; 66 (7): 726 – 735. Doi: https://doi.org/10.1016/j.jclinepi.2013.02.003

Appleby J, Devlin N (2004). Measuring Success in the NHS. Using patient-assessed health outcomes to manage the performance of health care providers. London: King's Fund / Dr Foster.

Assendelft, WJJ. Morton, SC. Yu, EI. Suttorp, MJ. Sheklle, PG. (2004). Spinal Manipulative Therapy for Low Back Pain. Cochrane Database Systematic Reviews; (1): CD000447. Doi: 10.1002/14651858.CD000447.pub2

Association of Chartered Physiotherapists in Occupational Health and Ergonomics (ACPOHE) (2015) Behaviours, knowledge and skills required by physiotherapists for working in Occupational Health. Version 5: http://www.acpohe.org.uk/sites/default/files/ACPOHE%20Competency%20Framework%202015%20v5.pdf

Atlas, SJ. (2017) Management of low back pain: getting from evidence-based recommendations to high-value care. Annals of Internal Medicine; 166 (7): 533 – 534. Doi: 10.7326/M17-0293

Ault, A. (2018) Specialists vs primary care: Whose guidelines matter? Medscape: https://www.medscape.com/viewarticle/898037

Balestroni, G. Bertolotti, G. (2012) EuroQol-5D (EQ-5D): an instrument for measuring quality of life. Monaldi Archives for Chest Disease; 78 (3): 155 – 159. https://doi.org/10.4081/monaldi.2012.121

Bassler, D. Montori, VM. Briel, M. Glasziou, P. Guyatt, G. (2007) Early stopping of randomized clinical trials for overt efficacy is problematic. Journal of Clinical Epidemiology; 61 (3): 241 – 246

Bausell, B. (2009) Acupuncture Research: Placebos by many other names. Archives of Internal Medicine; 169 (19): 1806 – 1818. Doi: 10.1001/archinternmed.2009.378

Beard, DJ. Rees, JL. Cook, JA. Rombach, I. Cooper, C. Merritt, N. Shirkey, BA. Donovan, JL. Gwilym, S. Savulescu, J. Moser, J. Gray, A. Jepson, M. Tracey, I. Judge, A. Wartolowska, K. Carr, AJ. (2018) Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicenter, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. The Lancet; 391 (10118): 329 – 338. Doi: https://doi.org/10.1016/S0140-6736(17)32457-1

Becker, A. Chenot, JF. Neibling, W. (2003) The German College of General Practitioners (DEGAM) low back pain guideline. Omikron Publishing. Düsseldorf, (in German)

Beecher (1995) The powerful placebo. Journal of the American Medical Association; 159 (17): 1602 – 1606. Doi: 10.1001/jama.1995.02960340022006

Bekkering, GE. Hendriks, BW. Koes, BW. Oostendorp, RAB. Ostelo, RWJG. Thomassen, JMC. Van Tulder, MW. (2003) National practice guideline for the physiotherapeutic management of patients with low back pain. Physiotherapy; 89 (2): 82 - 96

Belsey, J. (2002) Primary care workload in the management of chronic pain. A retrospective cohort study using a GP database to identify resource implications for UK primary care. Journal of Medical Economics, Journal of Medical Economics; 5: 39 – 50

Benedetti, F. Placebo and the new physiology of the doctor-patient relationship. (2013) Physiological Reviews; 93 (3): 1207 – 1246. Doi: 10.1152/physrev.00043.2012

Benedetti, F. Lanotte, M. Colloca, L. Ducati, A. Zibetti, M. Lopiano, L. (2009) Electrophysiological properties of thalamic, subthalamic and nigral

neurons during the anti-parkinsonian placebo response. The Journal of Physiology: Volume 587, Issue 15, pages 3869–3883

Benedetti, F. (2008) Mechanisms of Placebo and Placebo-Related Effects Across Diseases and Treatments. Annual Review of Pharmacology and Toxicology. Volume 48: 33-60. Doi: 10.1146/annurev.pharmtox.48.113006.094711

Bennell, K. Wee, E. Coburn, S. Green, S. Harris, A. Staples, M. Forbes, R. Buchbinder, R. (2010) Efficacy of standardised manual therapy and home exercise programme for chronic rotator cuff disease: randomised placebo controlled trial. BMJ; 340: c2756. Doi: 10.1136/bmj.c2756

Bernstein, I. Malik, Q. Carville, S. Ward, S. (2017) Low back pain and sciatic: Summary of NICE guidance. BMJ Clinical Research; 356: i6748. Doi: 10.1136/bmj.i6748

Beurskens, AJ. De Vet, HC. Koke, AJ. (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. Pain; 65: 71 - 76

Bialosky, JE. George, SZ. Horn, ME. Price, DD. Staud, R. Robinson, ME. (2014) Spinal manipulative therapy specific changes in pain sensitivity in individuals with low back pain. Journal of Pain; 15 (2): 136 – 148. Doi: 10.1016/j.jpain.2013.10.005

Bibby, AC. Torgerson, DJ. Leach, S. Lewis-White, H. Maskell, NA. (2018) Commentary: Considerations for using the Trials within Cohorts design in a clinical trial of an investigational medicinal product. Trials; 19 (18): 1 – 6. Doi: 10.1186/s13063-017-2432-3

Bijur, PE. Silver, W. Gallagher, EJ. (2001) Reliability of the visual analog scale for measurement of acute pain. Academic Emergency Medicine; 8 (12): 1153 -1157

Birch, S. Alraek, T. Lee, MS. (2016) Challenges for clinical practice guidelines in traditional medicines: The example of acupuncture.

European Journal of Integrative Medicine; 8 (4): 332 – 336. Doi: 10.1016/j.eujim.2016.07.032

Birch, S. Alraek, T. Robinson, N. Lee, MS. (2016) The U.K. NICE 2014 guidelines for Osteoarthritis of the knee: Lessons learned in a narrative review addressing inadvertent limitations and bias. The Journal of Alternative and Complimentary Medicine; 23 (4): 242 – 246. Doi: 10.1089/acm.2016.0385

Birch, S. Keppel Hesselink, J. Jonkman, FAM. Hekker, TAM. Bos, AAT. (2004). Clinical research on acupuncture: Part 1. What have reviews of the efficacy and safety of acupuncture told us so far? The Journal of Alternative and Complementary Medicine. 10, 468-480

Birch, S. Felt, R. (1999) Understanding Acupuncture. Edinburgh: Churchill Livingstone. UK

Bishop, FL. Lewis, G. Harris, S. McKay, N. Prentice, P. Thiel, H. Lewith, GT. (2010) A within-subject trial to test the equivalence of online and paper outcome measures: the Roland Morris Disability Questionnaire. BMC Musculoskeletal Disorders; 11 (113): Doi: 10.1186/1471-2474-11-113

Black, N. (2013) Patient reported outcome measures could help transform healthcare. *BMJ* 2013; 346:f167. Doi: http://dx.doi.org/10.1136/bmj.f167

Blackwood, B. (2006) Methodological issues in evaluation complex interventions. Journal of Advanced Nursing (JAN); 54 (5): 612 – 622. Doi: 10.1111/j.1365-2648.2006.03869.x

Blanchette, MC. Bussières, A. Stochkendahl, MJ. Boruff, J. Harrison, P. (2015) Effectiveness and economic evaluation of chiropractic care for the treatment of low back pain: a systematic review protocol. Systematic reviews; 4 (30)

Bland M. (2000) An Introduction to Medical Statistics, Third Edition Oxford University Press, Oxford, UK.

Bland JM. (2000b) Sample size in guideline trials. Family Practice; 17: S17 - S20

Blease, CR. Bishop, FL. Kaptchuk, TJ. (2017) Informed consent and clinical trials: where is the placebo effect? BMJ; 356. Doi: 10.1136/bmj.j463

Blease, C. (2012) The principle and parity: the 'placebo effect' and physician communication. Journal of Medical Ethics; 38 (4): 199 – 203. Doi: 10.1136/medethics-2011-100177

Bohman, T. Alfredsson, L. Jenson, I. Hallqvist, J. Vingård, E. Skillgate, E. (2014) Does a healthy lifestyle behaviour influence the prognosis of low back pain among men and women in a general population? A population based cohort study. BMJ Open; 4 (12): http://dx.doi.org/10.1136/bmjopen-2014-005713

Bombardier, C. Esmail, R. Nachemson, AL. (1997) The Cochrane Collaboration Back Review Group for spinal disorders. Spine; 22: 837 – 840

Bondemark, L. Ruf, S. (2015) Randomized controlled trials: the gold standard or an unobtainable fallacy? European Journal of Orthodontics; 37 (5): 457 – 461. Doi: 10.1093/ejo/cjv046

Boonstra, AM. Schiphorst, PHR. Reneman, MF. Posthumus, JB. Stewart, RE. (2008) Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. International Journal of Rehabilitation Research; 31 (2): 165 – 169. Doi: 10.1097/MMR.0b013e3282fc0f93

Bouwmeester, W. van Enst, A. van Tulder, M. (2009) Quality of low back pain guidelines improved. Spine; 34 (23): 2562 - 2567. Doi: 10.1097/BRS.0b013e3181b4d50d

Bovey, M. (2017) Acupuncture: is NICE cutting off its nose to spite its face? A response to: Low back pain and sciatica: summary of NICE

guidance. BMJ; 365: i6748 https://www.bmj.com/content/356/bmj.i6748/rr-7

Bowen Hills (2003) Australian acute musculoskeletal pain guidelines. Evidence-based management of acute musculoskeletal pain. Bowen Hills: Australian Academic Press

Bowling, A. Ebrahim, S. (2005) Handbook of health research methods, investigation, measurement and analysis. Open University Press

Boyd, A. Tilling, K. Cornish, R. Davies, A. Humphries, K. Macleod, J. (2015) Professionally designed information material and telephone reminders improved consent response rates: evidence from an RCT nested within a cohort study. Journal of Clinical Epidemiology; (68) 8: 877 – 887. Doi: 10.1016/j.jclineoi.2015.03.014

Bradley-Gilbride, J. Bradley C. (2010) Partially randomised preference trial design. In Neil J. Salkind (Ed) Encyclopedia of Research Design. USA: Sage; volume 2: 1009 - 1015

Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. (2006) Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. European Journal of Pain; 10 (4): 287-333

Brewin, CR. Bradley, C. (1989) Patient preferences and randomised clinical trials. British Medical Journal; 299: 313 – 15

Bridges S. (2012) Chronic pain. In: Craig, R. Mindell, J. (2012) Health Survey for England 2011 Health, social care and lifestyles.356. Health and Social Care Information Centre: 1 – 33

Brinjikji, W. Luetmer, B. Comsock, BW. Bresnahan, LE. Chen, LE. Deyo, RA. Halabi, S. Turner, JA. Avins, AL. James, K. Wald, JT. Kallmes, DF. Jarvik, JG. (2015) Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. American Journal of Neurology; 36 (4): 811 – 816. Doi: 10.3174/ajnr.A4173

Brinkhaus B1, Witt CM, Jena S, Liecker B, Wegscheider K, Willich SN. (2008) Acupuncture in patients with allergic rhinitis: a pragmatic randomized trial. Annuals of Allergies, Asthma and Immunology. 101(5): 535-43. Doi: 10.1016/S1081-1206 (10) 60294-3

Brinkhaus, B. Witt, CM. Jena, S. Linde, K. Streng, A. Wagnepfeil, S. Irnich, D. Walther, HU. Melchart, D. Willich, SN. (2006) Acupuncture in patients with chronic low back pain: a randomized controlled trial. JAMA Internal Medicine; 166 (4): 450 – 457. Doi: 10.1001/archinte.166.4.450

Brosseau, L. Wells, GA. Poitras, S. Tugwell, P. Casimiro, L. Novikov, M. Loew, L. Sredic, D. Clément, S. Gravelle, A. Kresic, D. Hua, K. Lakic, A. Ménard, G. Sabourin, S. Bolduc, MA. Ratté, I. McEwan, J. Furlan, A. Gross A. Dagenais, S. Dryden, T. Muckenheim, R. Côté, R. Paré, V. Rouhani, A. Léonard, G. Finestone, HM. Laferrière, L. Haines-Wangda, A. Russell-Doreleyers, M. De Angelis, G. Cohoon, C. (2012) Ottawa panel evidence-based clinical practice guidelines on therapeutic massage for low back pain. Journal of bodywork and movement therapies. 16 (4): 424 -455. Doi: 10.1016/j.jbmt.2012.04.002

Brouwer, S. Kuijer, W. Dijkstra, PU. Göeken, LNH. Groothoff, JW. Geertzen, JHB. (2004) Reliability and stability of the Roland Morris Disability Questionnaire: Intra Class Correlation and limits of agreement. Disability and Rehabilitation; 26 (3): 162 – 165. Doi: 10.10.80/09638280310001639713

Burton, K. Main, C. Cantrall, T. (2002). The Back Book, 2nd Edition - UK Edition, London, TSO (The Stationery Office)

Buchbinder, R. van Tulder, M. Öberg, B. Costa, LM. Woolf, A. Schoene, M. Croft, P. (2018) Low back pain: a call for action. The Lancet; 391: 2384 – 2388. Doi: 10.1016/S0140-6736(18)30488-4

Bradnam, LV. (2011) A bio psychosocial clinical reasoning model for western acupuncture. Physical Therapy Reviews; 16 (2): 138 – 146. Doi: 10.1179/1743288X11Y.0000000011

BPS / British Pain Society (2019) Outcome Measures. Faculty of Pain Medicine of the Royal College of Anaesthetists. January 2019. https://www.britishpainsociety.org/static/uploads/resources/files/Outcome_Measures_January_2019.pdf

Broer, T. Bal, R. Pickersgill, M. (2016) Problematisations of complexity: On the notion and production of diverse complexities in healthcare interventions and evaluations. Science as Culture; 26 (2): 135 – 160. Doi: 10.1080/09505431.2016.1212003

Brunner, E. De Herdt, A. Minguet, P. Baldew, SS. Probst, M. (2013) Can cognitive behavioural therapy based strategies be integrated into physiotherapy for the prevention of chronic low back pain? A systematic review. Disability Rehabilitation; 35 (1): 1 – 10. Doi: 10.3109/09638288.2012.683848

Bystad, M. Bystad C. Wynn, R. (2015) How can placebo effects best be applied in clinical practice? A narrative review. Psychology Research and Behaviour Management; 8: 41 – 45. Doi: 10.2147/PRBM.S75670

Byth, K. Gebski, V. (2004) Factorial designs: a graphical aid for choosing study designs accounting for interaction. Clinical Trials. 1; 315-325

Cambron, JA. Jedlicka, J. Iannelli, GC. Dexheimer, J. Cramer, GD. (2011) Patients perception of Care: Comparing Placebo and Manual Therapy Within a Pilot Study. Topics in Integrative Health Care; Vol. 2(2) ID: 2.2003

Campbell, A. (2006) Point specificity of acupuncture in the light of recent clinical imaging studies. Acupuncture in Medicine; 24 (3): 118 – 122. Doi: 10.1136/aim.24.3.118

Campbell, R. Peters, T. Grant, C. Quilty, B. Dieppe, P. (2005) Adapting the randomized consent (Zelen) design for trials of behavioural interventions for chronic disease: feasibility study. Journal of Health Services Research & Policy; 10 (4): 220 - 225

Campbell, M. Fitzpatrick, R. Haines, A. Kinmouth, AL. Sandercock, P. Spiegelhalter, D. Tyrer, P. (2000) Framework for design and evaluation of complex interventions to improve health. BMJ 321: 694. Doi: https://doi.org?10.1136/bmj.321.7262.694

Caporaso, F. Pulkovski, N. Sprott, H. Mannion, AF. (2012) How well do observed functional limitations explain the variance in Roland Morris scores in patients with chronic non-specific low back pain undergoing physiotherapy. European Spine; 21 (2): S187 – 195. Doi: 10.1007/s00586-012-2255-6

Care Quality Commission (CQC) (2017) The state of health care and adult social care in England 2016 /17. Parliamentary document 83 (4) (a) of the Health and Social Care Act. ID. CCS0917121928 10/17 https://www.cqc.org.uk/sites/default/files/20171010_stateofcare1617_report.pdf

Carrasco-Labra, A. Brignardello-Petersen, R. Santesso, N. Neumann, I. Mustafa, RA. Mbuagbaw, L. Ikobaltzeta, IE. De Sito, C. McCullagh, LJ. Alonso-Coello, P. Meerpohl, JJ. Vandvik, PO. Brozek, JL. Akl, EA. Bossuyt, P. Churchill, R. Glenton, C. Rosenbaum, S. Tugwell, P. Welch, V. Guyatt, G. Schünemann, H. (2015) Comparison between the standard and a new alternative format of the Summary-of-Findings tables in Cochrane review users: study protocol for a randomized controlled trial. Trials; 16: 164. Doi: 10.1186/s13063-015-0649-6

Chan, AW. Altman, DG. (2005) Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. BMJ. 330 (7494): 753. Doi: 10.1136/bmj.38356.424606.8F

CSP – Chartered Society of Physiotherapy (CSP) (2017) 10 things you need to know about your back. http://www.csp.org.uk/yourback

CSP – Chartered Society of Physiotherapy (CSP) (2014) Report, May - http://www.csp.org.uk/professional-union/practice/your-business/evidence-base/outcome-experience-measurement/eq-5d-5l

CSP – Chartered Society of Physiotherapy (CSP) (2014a) CSP annual report. http://www.csp.org.uk/about-csp/what-we-do/annual-report

CSP – Chartered Society of Physiotherapy (CSP) (2012) Landmark decision gives UK physios a world first in prescribing rights. http://www.csp.org.uk/news/2012/07/23/landmark-decision-gives-uk-physios-world-first-prescribing-rights

CSP – Chartered Society of Physiotherapy (CSP) (2007) Guidelines for the management of persistent low back pain. London; The Chartered Society of Physiotherapy: Part: 1 and 2

Chen, YJ. Chen, CT. Liu, JY. Bassi, GS. Yang, YQ. (2019) What is the appropriate acupuncture treatment schedule for chronic pain? Review and analysis of randomized controlled trials. Evidence-Based Complementary and Alternative Medicine. Doi: https://doi.org/10.1155/2019/5281039

Chen, Z. Li, Y. Zhang, X. Chen, S. Yang, W. Zheng, X. Zheng, G. (2017) Sham electro-acupuncture methods in randomized controlled trials. Scientific Reports; 7: 40837. Doi: 10.1038/srep40837

Cherkin, DC. Sherman, KJ. Avins, AL. Erro, JH. Ichikawa, L. Barlow, W. Delaney, K. Hawkes, R. Hamilton, L. Pressman, A. Khalso, PS. Deyo, RA. (2009) A randomised controlled trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. Archives of Internal Medicine; 169 (9): 858 - 866

Cherkin, DC. Eisenberg, D. Sherman, KJ. Barlow, W. Kaptchuk, TJ. Street, J. Deyo, RA. (2001) Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage and self-care education for chronic low back pain. Archives Internal Medicine; 161: 1081

Cherkin, DC. Deyo, RA. Battié, M. Street, J. Barlow, W. (1998) A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. New England Journal of Medicine; 339 (15): 1021 – 1029. Doi:

10.1056/NEJM199810083391502

Chiarotto, A. Boers, RA. Deyo, RA. Buchbinder, R. Corbin, TP. Costa, LOP. Foster, NE. Grotle, M. Koes, BW. Kovacs, FM. Lin, CW. Maher, CG. Pearson, AM. Peul, WC. Schoene, ML. Turk, DC. Van Tulder, MW. Terwee, CB. Ostelo, RW. (2018) A core outcome measurement set for low back pain clinical trials. Orthopaedic Proceedings; 110-B (2): 27th February. https://online.boneandjoint.org.uk/doi/abs/10.1302/1358-992X.2018.2.029

Chiarotto, A. Maxwell, LJ. Terwee, CB. Wells, GA. Tugwell, P. Ostelo, RW. (2016) Roland – Morris Disability Questionnaire and Oswestry Disability Index: Which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and Meta-Analysis. Physical Therapy; 96 (10): 1620 – 1637. Doi: 10.2522/ptj.20150420

Cho, YJ. Song, YK. Cha, YY. Shin, BC. Shin, IH. Park, HJ. Lee, HS. Kim, KW. Cho, JH. Chung, WS. Lee, JH. Song, MY. (2013) Acupuncture for chronic low back pain: a multicentre, randomized, patient-assessor blind, sham-controlled clinical trial. Spine; 38 (7): 549 - 557

Choi EM. Fang, J. Longhurst, JC. (2012) Point specificity in acupuncture. Chinese Medicine; 7: 4 http://www.cmjournal.org/content/7/1/4

Cholangitis, PS. (2018) Low back pain – Executive summary. The Lancet; 391: (10137)

Chong, OT. Critchley, HOD. Horne, AW. Elton, R. Haraldsdottir, E. Fallon, M. (2015) The BMEA study: the impact of meridian balanced method electro-acupuncture on women with chronic pelvic pain – a three – arm randomised controlled pilot study using mixed – methods approach. BMJ open; 5 (11): 1 – 9. Doi: 10.1136/bmjopen-2015-008621

Chou R, Chekelle P. (2010) Will this patient develop persistent disabling low back pain? JAMA; 303: 1295 – 302

Chou, R. Loeser, JD. Owens, DK. Rosenquist, RW. Atlas, SJ. Baisden, J. Carragee, EJ. Grabois, M. Murphy, DR. Resnick, DK. Stanos, SP. Shaffer, WO. Wall, EM. For the American Pain Society Low Back Pain Guideline Panel (2009) Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain. An evidence-based clinical practice guideline from the American Pain Society. SPINE; 34 (10): 1066 - 1077

Chou, R. Qaseem, A. Snow, V. Casey, D. Cross, JT. Shekelle, P. Owens, DK. Clinical Efficacy Assessment Subcommittee of the American College of Physicians; American College of Physicians; American Pain Society Low Back Pain Guidelines Panel. (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; October 2; 147 (7) 478 – 491

Chou, R. Huffman, LH. (2007). Nonpharmacological therapies for acute and chronic Low Back Pain: A review of the evidence for an American Pain Society/American college of Physicians Clinical Practice Guideline. Annals of Internal Medicine; 147: 492 – 504

Chow, SC. Chang, M. (2008) Adaptive designs methods in clinical trials – a review. Orphanet Journal of Rare Diseases; 3 (11): Doi: 10.1186/1750-1172-3-11

Clark, S. Horton, R. (2018) Low back pain: a major global challenge. The Lancet; March 21: https://doi.org/10.1016/S0140-6736(18)30725-6

Claxton, K. Sculpher, M. Drummond, M. (2002) A rational framework for decision making by the National Institute for Clinical Excellence (NICE) (2002) The Lancet; 360 (9334) 711 – 715. Doi: 10.1016/S0140-6736(02)09832-X

Clinical knowledge summaries (CKS) National Institute for Clinical Excellence (NICE) Back pain (low) and sciatica. www.cks.library.nhs.uk. Accessed Sept 2008

Coe, R. (2002) It's the effect size, stupid. What effect size is and why is it important. https://www.leeds.ac.uk/educol/documents/00002182.htm

Cochrane (2018) Assessing risk of bias in included studies. Cochrane Methods Bias; https://methods.cochrane.org/bias/assessing-risk-bias-included-studies

Cockayne, S. Fairhurst, C. Adamson, Hewitt, C. Hull, R. Hicks, K. Keenen, AM. Lamb, SE. Green, L. McIntosh, C. Menz, HB. Redmond, AC. Rodgers, S.J. Torgerson, DJ. Vernon, W. Watson, J. Knapp, P. Rick, J. Bower, P. Eldridge, S. Madurasinghe, VW. Graffy, J. (2017) An optimised patient information sheet did not significantly increase recruitment or retention in a falls prevention study: an embedded randomised recruitment trail. Trials; 18: 144. Doi: 10.1186/s13063-017-1797-7

Cohen, J. (1992) A power primer. Psychological Bulletin, Volume 112 (1), July 155 - 159 http://dx.doi.org/10.1037/0033-2909.112.1.155

Colquhoun, D. Novella, SP. (2013) Acupuncture is theatrical placebo. Journal of Anaesthesia & Analgesia; 116 (6): 1360 – 1363. http://www.dcscience.net/Colquhoun-Novella-A&A-2013.pdf

COST ACTION B13 (2006) COST ACTION B13: European guidelines for the management of low back pain. European Spine Journal; 15 (2)

Costa, LCM. Maher, CG. McAuley, JH. Hancock, MJ. Herbert, RD. Refshauge, KM. Henschke, N. (2009) Prognosis for patients with chronic low back pain: inception cohort study. BMJ; Oct 6; 339: b3829. Doi: 10.1136/bmj.b3829

Couper, MP. Tourangeau, R. Conrad, FG. (2006) Evaluating the effectiveness of visual analog scales: A web experiment. Social Science Computer Review; 24; 227 - 245. https://ssc.sagepub.com/cgi/content/abstract/24/2/227

Cox, H. Tilbrook, H. Aplin, J. Chuang, LH. Hewitt, C. Jayakody, S.

Semlyen, A. Soares, MO. Torgerson, D. Trewhela, A. Watt, I. Worthy, G. (2010). A pragmatic multi-centred randomised controlled trial of yoga for chronic low back pain: Trial protocol. Complementary Therapies in Clinical Practice; 16 (2): 76 – 80. Doi: 10.1016/j.ctcp.2009.09.010

Cutforth, G. Peter, A. Taenzer, P. (2011) The Alberta Health Technology Assessment (HTA) Ambassador program: The development of a contextually relevant, multidisciplinary clinical practice guideline for non specific low back pain: a review. Physiotherapy Canada; 63 (3): 278 – 286. Doi: 10/3138/ptc.2009-39P

Cummings, M. Hróbjartsson, A. Ernst, E. (2018) Should doctors recommend acupuncture for pain? BMJ; 360: K970. Doi: 10.1136/bmj.k970

Cummings, M. (2017) Response to Low back pain and sciatica: summary of NICE guidance. BMJ; 356: i6748. Doi: 10.1136/bmj.i6748

Cummings, M. (2016) Quality sham – there appears to be a significantly greater improvement in health related quality of life (HRQoL) with sham acupuncture than with conventional care. Acupuncture in Medicine; BMJ Blog October 13: https://blogs.bmj.com/aim/2016/10/13/quality-sham/

Cummings (2016b) Too NICE – there appears to be a glaring orthodox bias in NG59. Acupuncture in Medicine; BMJ Blog November 30: https://blogs.bmj.com/aim/2016/11/30/too-nice/

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. British Medical Journal (BMJ); volume 337

Critchley, DJ. Ratcliffe, J. Noonan, S. Jones, RH. Hurley, MV. (2007) Effectiveness and cost effectiveness of three types of physiotherapy used to reduce chronic low back pain disability: a pragmatised randomised trial with economic evaluation. Spine; 32 (14): 1474 – 1481

Crow, WH. Willis, DR. (2009) Estimating cost of care for patients with

acute low back pain: A retrospective review of patient records. The Journal of the American Osteopathic Association; 109: 229 - 233

Czarnawska-Iliev, I. Robinson, N. (2016) General Practitioners' use and attitudes to acupuncture in relation to UK's National Institute for Health and Care Excellence (NICE) clinical guidelines – a pilot study. European Journal of Integrative Medicine. http://dx.doi.org/10.1016/j.eujim.2016.07.004

D'Agostino, RB. (2009) The delayed-start study design. New England Journal of Medicine; 361: 1304 – 1306

Da Silva, MD. Bobinski, F. Sato, KL. Kolker, SJ. Sluka, KA. Santos, ARS. (2014) IL-10 cytokine released from M2 macrophages is crucial for analgesic and anti-inflammatory effects of acupuncture in a model of inflammatory muscle pain. Molecular Neurobiology; 51 (1): 19 31. Doi: 10.1007/s12035-014-8790-x

Dagenais, S. Caro, J. Haldeman, S. (2008) A systematic review of low back pain cost of illness studies in the United States and internationally Spine; 8: 8 – 20

Darlow, B. Dean, S. Perry, M. Mattieson, F. Baxter, D. Dowell, A. (2014) Acute low back pain management in general practice: uncertainty and conflicting certainties. Family Practice; 31 (6): 723 - 732

Dascanio, VC. (2015a) The Acupuncture Association of Chartered Physiotherapists, Chairman's report. Acupuncture in Physiotherapy; 27: 2

Dascanio, VC. (2015b) Acupuncture in Physiotherapy: a contemporary UK perspective. Acupuncture in Medicine; 33: 442 – 444

Dascanio, VC. (2015c) A pilot factorial randomised cohort trial of manual therapy or acupuncture for low back pain. Conference presentation. British Medical Acupuncture Society (BMAS) Spring Scientific Meeting, Newcastle-upon-Tyne

Dascanio, VC. (2015d) A pragmatic pilot factorial randomized controlled trial of acupuncture versus manual therapy for low back pain nested within an observational cohort study: a trial protocol. Acupuncture in Physiotherapy; 27 (2): 19 – 36

Dascanio, V. van Dongen, C. Ireland, CM. Wilson, AJ. (2015e) Acupuncture in Physiotherapy. The Evidence: A summary of evidence for the use of acupuncture in physiotherapy for the benefit of the patient. Publication of: the Acupuncture Association of Chartered Physiotherapists Limited (AACP Ltd) https://www.aacp.org.uk/assets/ckfinder_library/files/150527%20AACP%20The%20Evidence%20(1).pdf

Dascanio, V. van Dongen, C. Ireland, CM. Wilson, AJ. (2015f) Acupuncture in Physiotherapy: Commissioning Guidance: Guidance on commissioning for AACP members. Publication of: the Acupuncture Association of Chartered Physiotherapists Limited (AACP Ltd) https://www.aacp.org.uk/assets/ckfinder_library/files/150527%20AACP%20Commissioning%20Guidance%20.pdf

Dascanio, V, Birks, Y, Clark, LK, Fairhurst, C, MacPherson, H & Torgerson, DJ (2014) Randomized cohort trial was shown to be feasible for evaluating treatments in low back pain. Journal of Clinical Epidemiology, 67 (8): 940 – 946 http://dx.doi:10.1016/j.jclinepi.2014.04.004

Dascanio, V. Birks, Y. Torgerson, D. (2011) A pilot factorial randomised cohort trial of manual therapy or acupuncture for low back pain – Conference abstract and poster presentation. Trials; 12 (1): A150

Data Protection Act (1998) UK Act of Parliament. TSO (The Stationary Office); (29) London, UK

Davidson, M. Keating, JL. (2002) A Comparison of Five Low Back Disability Questionnaires: Reliability and Responsiveness. Physical Therapy; 82: 8 – 24

Davis, N. (2016) Acupuncture for low back pain no longer recommended for NHS patients. The Guardian; UK edition; Science News: 24.3.2016. https://www.theguardian.com/science/2016/mar/24/acupuncture-for-low-back-pain-no-longer-recommended-for-nhs-patients

Deakin, PJ. Tsai, J. (2017) Who buys herbal medicines and what products are they using? Secondary analysis of a cross-sectional health database in Yorkshire, UK. Herbal Medicine, UK

De Hertogh, W. Vaes, P. Devroey, D. Louis, P. Carpay, H. Truijen, S. Duquet, W. Oostendorp, R. (2009) Preliminary results, methodological considerations and recruitment difficulties of a randomised clinical trial comparing two treatment regimes for patients with headache and neck pain. BMC Musculoskeletal Disorders; 10: 115. Doi: 10.1186/1471-2472

De Jong, RH. (1995) Backfire: AHCPR practice guideline for acute low back pain. Journal of South Carolina Medical Association; 91 (11): 465 468

Delgado, DA. Lambert, BS. Boutris, N. McCulloch, PC. Robbins, AB. Moreno, MR. Harris, JD. (2018). Validation of digital visual analog scale pain scoring with a traditional paper-based visual analogue scale in adults. Journal of the American Academy of Orthopaedic Surgeons; 2 (3): e088. Doi: 10.5435/JAAOSGlobal-D-17-00088

Delitto, A. George, SZ. Van Dillen, L. Whitman, JM. Sowa, G. Shekelle, P. Denninger, TR. Godges, JJ. (2012) Low back pain: Clinical practice guidelines linked to the international classification of functioning, disability, and health from the orthopaedic section of the American physical therapy association. Journal of Orthopaedic & Sports Physical Therapy; 42 (4): A1 – A57. Doi: 10.2519/jospt.2012.0301

Della-Giustina, D. (2015) Evaluation and treatment of acute back pain in the emergency department. Emergency Medicine Clinics of North America; 33 (2): 311 – 326 https://doi.org/10.1016/j.emc.2014.12.005

Department of Health (DOH) Health (2006) The Musculoskeletal Services framework. A Joint Responsibility: doing it differently. Gateway ref: 6857, http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4138412.pdf

Department of Health (DOH) (2001) Research governance framework for health and social care. London: Department of Health Publications. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/139565/dh_4122427.pdf

Department of Health (DOH) (1999) Statistics Division. The prevalence of back pain in Great Britain in 1998. London: Government Statistical Service

Department of Work and Pensions (2003) Cost of LBP. London: Department of Work and Pensions publication.

Deyo, RA. Dworkin, SF. Amtmann, D. Andersson, G. Borenstein, D. Carragee, E. Carrino, J. Chou, R. Cook, K. DeLitto, A. Goertz, C. Khalsa, P. Loeser, J. Mackey, S. Panagis, J. Rainville, J. Tosteson, T. Turk, D. Von Korff, M. Weiner, DK. (2014) Report of the NIH Task Force on research standards for chronic low back pain. Spine (Phila Pa 1976). 2014; 39 (14): 1128 - 1143 Doi: 10.1097/brs.0000000000000434

Deyo, RA. Weinstein, JN. (2001) Low back pain. New England Journal of Medicine. 344 (5): 363 – 370

Deyo, RA. Battie, M. Beurskens, AJ. Bombardier, C. Croft, P. Koes, B. Malmivaara, A. Roland, M. Von Korff, M. Waddell, G. (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine; 23 (18): 2003 -2013

Dionne CE. (1999) Low back pain. In Crombie IK, Croft PR, Linton SJ, LeResche L, Von Korff M, eds. Epidemiology of pain: a report of the task force on epidemiology of the International Association for the Study of

Pain, Seattle, WA: IASP Press, 1999

Doran, DM. Newell, DJ. (1975) Manipulation in treatment of low back pain: a multicentre study. British Medical Journal (BMJ); 2 (5964): 161 - 164

Drug Committee of the German Medical Society (2007) Recommendations for treatment of low back pain. Köln, Germany

Dufour, N. Thamsborg, G. Oefeldt, A. Lundsgaard, C. Stender, S. (2010) Treatment of Chronic Low Back Pain: A Randomized, Clinical Trial Comparing Group-Based Multidisciplinary Biopsychosocial Rehabilitation and Intensive Individual Therapist-Assisted Back Muscle Strengthening Exercises. Spine; 35 (5): 469 - 476

Dumville, JC. Torgerson, DJ. Hewitt, CE. (2006) Reporting attrition in randomised controlled trials. British Medical Journal (BMJ); volume 332 (7547) 969 – 971

Dutch Institute for Healthcare Improvement (CBO) (2003) Clinical guideline for non-specific low back pain. Dutch Institute for Healthcare Improvement (CBO)

Duthey, B. (2013) Background paper 6.24 Low Back Pain. Priority Medicines for Europe and the World "A Public Health Approach to Innovation"

Edelist, G. Gross, AE. Langer, F. (1976) Treatment of low back pain with acupuncture. Canadian Anaesthetists Society Journal; 23 (3): 303 - 306

Edwards, J. Hayden, J. Asbridge, M. Gregoire, B and Magee, K. (2016) Prevalence of low back pain in emergency settings: a systematic review and meta-analysis. BMC Musculoskeletal Disorders; 18: 143 https://doi.org/10.1186/s12891-017-1511-7

Edwards, P. Clarke, M. DiGuiseppi, C. Pratap, S. Roberts, I. Wentz, R. (2002) Identification of randomized controlled trials in systematic

reviews: accuracy and reliability of screening records. Statistics in Medicine; 21: 1635 – 1640

Ekman, M. Jönhagen, S. Hunsche, E. Jönsson, L. (2005) Burden of illness of chronic low back pain in Sweden: a cross-sectional, retrospective study in primary care setting. Spine; 30 (15): 1777 - 1785

Eldridge, SM. Lancaster, GA. Campbell, MJ. Thanbane, L. Hopewell, S. Coleman, CL. Bond, CM. (2016) Defining Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of a Conceptual Framework. PLOS ONE 11(3): http://dx.doi.org/10.1371/journal.pone.0150205

Ellard, DR. Underwood, M. Achana, F. Antrobus, JHL. Balasubramanian, S. Brown, S. Cairns, M. Griffin, J. Griffiths, F. Haywood, K. Hutchinson, C. Lall, R. Petrou, S. Stallard, N. Tysall, C. Walsh, DA. Sandhu, H. (2017) Interpreting treatment effects: 'What is the minimal between-group difference in change scores necessary for facet joint injection to be considered worthwhile?' Health Technology Assessment; 21.30

Ernst, E. (2016) Data fabrication in China is an open secret. Blog post. https://edzardernst.com/2016/10/data-fabrication-in-china-is-an-open-secret/

Ernst, E. Lee, MS. Choi, TY. (2010) Acupuncture: does it alleviate pain and are there serious risks? A review of reviews. Pain: 152: 755 – 764. Doi: 10.1016/J.pain.2010.11.004

Ernst, E. (2008) Acupuncture ineffective, attention effective? Archives of Internal Medicine; 168 (5): 551. Doi: 10.1001/archinte.168.5.551

Ernst, E. White, AR. (1998). "Acupuncture for back pain: a meta-analysis of randomized controlled trials." Archives International Medicine; 158 (20): 2235 – 2241

Evans, R. Haas, M. Schulz, C. Leininger, B. Hanson, L. Bronfort, G. (2018) Spinal manipulation and exercise for low back pain in adolescents:

a randomized trial. Pain; 159 (7): 1297 – 1307. Doi: 10.1097/j.pain0000000000001211

Fairbank, JCT. (2014) Why are their different versions of the Oswestry Disability Index? A review. Journal of Neurosurgery Spine; 20: 83 – 86. Doi: 10.3171/2013.9.SPINE13344

Fairbank, JC. Pynsent, PB. (2000) The Oswestry Disability Index. Spine; 25 (22): 2940 - 2952

Fairbank, JC. Couper, J. Davies, JB. (1980). The Oswestry Low Back Pain Disability Questionnaire. Physiotherapy; 66: 271 - 273

Falagas, ME. Grigori, T. Loannidou, E (2009) A systematic review of trends in the methodological quality of randomized controlled trials in various research fields. Journal of Clinical Epidemiology; 62 (3): 227 - 231 Doi: 10.1016/j.clinepi.2008.07.012

Finlay, A. McAlistar, FA. Straus, SE. Sackett, DL. Altman, DG. (2003) Analysis and reporting of factorial trials: a systematic review. JAMA; 289: 2545 – 2553

Fischer, F. Lange, K. Greiner, W. Kraemer, A. (2016) Barriers and strategies in guideline implementation – A scoping review. Healthcare; 4 (3): 36 Doi: 10.3390/healthcare4030036

FDA – Food and Drug Administration (FDA) (2009) Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. http://wwwfdagov/downloads/Drugs/GuidanceCompliance RegulatoryInformation/Guidances/UCM193282pdf

Ford, I. Norrie, J. (2016) Pragmatic Trials. The New England Journal of Medicine; 375: 454 – 463 doi: 10.1056/NEjMral1510059

Forward, DP. Wallace, WA. Hughes, SPF. (2008) Synopsis of causation; low back pain. Ministry of Defence. Government Report

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384519/low_back_pain.pdf

Foster, NE. Anema, JR. Cherkin, D. Chou, R. Cohen, SP. Gross, DP. Ferreira, PH. Fritz, JM. Koes, BW. Peul, W. Turner, JA. Maher, CG. Lancet Low Back Pain Series Working Group. (2018) Pain and prevention of low back pain: evidence, challenges, and promising directions. Lancet; 9:391(10137): 2368 – 2383. Doi 10.1016/S0140-6736(18)30489-6

Foster, NE. Hill, JC. O'Sullivan, P. Hancock, M. (2013) Stratified models of care. Best Practice and Research: Clinical Rheumatology. 27 (5): 649 - 661

Foster, NE. Mullis, R. Hill, JC. Lewis, M. Whitehurst, DGT. Doyle, C. Konstantinou, K. Main, C. Somerville, S. Sowden, G. Wathall, S. Young, J. Hay, EM. IMPaCT Back Study team. (2014) Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. Annals of Family Medicine. 12 (2): 102 – 111

Frieden, TR. (2017) Evidence for health decision making – beyond randomized controlled trials. The New England Journal of Medicine; 377: 465 – 475 Doi: 10.1056/NEjMra1614394

Friedrich, M. Likar, R. (2007) Evidenz- und konsensusbasierte österreichische Leitlinien für das Management akuter und chronischer unspezifischer Kreuzschmerzen. Wien Klin Wochenschr; 119 (5 – 6): 189 – 197. Doi: 10.1007/s00508-006-0754-3

Fritz, JM. Childs, JD. Wainner, RS. Flynn, TW. (2012) Primary care referral of patients with low back pain to physical therapy. Spine; 37 (25): 2114 – 2121. Doi: 10.1097/BRS.0b013e31825d32f5

Fritz, JM. Irrgang, JJ. (2001) A comparison of a Modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Physical Therapy; 81 (2): 776 – 788. Doi: 10.1093/ptj/81.2.776

Frost, H. Lamb, SE. Stewart-Brown, S. (2008) Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index v2.1 and Roland and Morris Disability Questionnaire for Patients with sub acute and chronic low back pain. Spine; 33 (22): 2450 – 2457. Doi: 10.1097/BRS.0b013e31818916fd

Frymoyer, JW. (1988) 'Medical progress: Back pain and sciatica'. New England Journal of Medicine; 318: 291 – 300. Doi: 10.1056/NEJM198802043180506

Fullen, BM. Maher, T. Bury, G. Tynan, A. Daly, LE. Hurley, DA. (2007) Adherence of Irish general practitioners to European guidelines for acute low back pain: a prospective pilot study. European Journal of Pain; 11 (6) 614 – 623. Doi: 10.1016/j.ejpain.2006.09.007

Furlan, AD. Giraldo, M. Baskwill, A. Irvin, E. Imamura, M. (2015) Massage for low back pain. Cochrane Databases for Systematic Reviews; 1 (9) Doi: 10.1002/14651858.CD001929.pub3

Furlan, AD. Malmivaara, A. Chou, R. Maher, CG. Deyo, RA. Schoene, M. Bronfort, G. van Tulder, MW. Editorial Board of the Cochrane Back and Neck Group. (2015) 2015 updated methods guideline for systematic reviews in the Cochrane back and neck group. Spine; 40 (21): 1660 – 1673. Doi: 10.1097/BRS.0000000000001061

Furlan, A. Yazdi, F. Tsertsvadze, A. Gross, A. Van Tulder, M. Santaguida, L. Gagnier, J. Ammendolia, C. Dryden, T. Doucette, S. Skidmore, B. Daniel, R. Ostermann, T. Tsouros, S. (2012) A Systematic Review and Meta-Analysis of Efficacy, Cost-Effectiveness, and Safety of Selected Complementary and Alternative Medicine for Neck and Low-Back Pain. Evidence Based Complementary and Alternative Medicine; 2012: 953139. Published online 2011 Nov 24. Doi: 10.1155/2012/953139

Furlan, A. Yazdi, F. Tsertsvadze, A. Lao, L. Sherman, K. Ammendolia, C. Wang, E. Shin, B. Wakayama, I. van Tulder, M. (2011) Acupuncture for

(sub) acute non-specific low-back pain. Editorial Group: Cochrane Back Group protocol – Cochrane library. Published Online: 10 August

Furlan, A. Yazdi, F. Tsertsvadze, A. Gross, A. Van Tulder, M. Santaguida, L. Cherkin, D. Gagnier, J. Ammendolia, C. Ansari, M. Ostermann, T. Dryden, T. Doucette, S. Skidmore, B. Daniel, R. Tsouros, S. Weeks, L. Galipeau, J. (2010) Complementary and Alternative Therapies for Back Pain II. Evidence Report/Technology Assessment No. 194. (Prepared by the University of Ottawa Evidence-based Practice Center under Contract No. 290-2007-10059-I (EPCIII). AHRQ Publication No. 10(11) E007. Rockville, MD: Agency for Healthcare Research and Quality

Furlan, A. Pennick, V. Bombardier, C. van Tulder, M. Editorial Board Cochrane Back Review Group (2009) Updated method guidelines for systematic reviews in the Cochrane back review group. Spine. Aug 15; 34 (18): 1929 – 1941. Doi: 10.1097/BRS.0b013e3181b1c99f

Furlan, A. Imamura, M. Dryden, T. Irvin, E. (2008) Massage for low-back pain. Cochrane Database Systematic Reviews. October 8; (4): CD001929. Doi: 10.1002/14651858.CD001929.pub2

Furlan, A. van Tulder, MW. Cherkin, D. Tsukayama, H. Lao, L. Koes, BW. Berman, BM. (2005). Acupuncture and dry-needling for low back pain. Cochrane Database Systematic Review: CD001351

Furlan, AD. Brosseau, L. Imamura, M. Irvin, E. (2002) Massage for low-back pain: a systematic review within the framework of the Cochrane collaboration back review group. Spine; 27 (17): 1896 – 1910

Frank, JW. Brooker, AS. DeMaio, SE. Kerr, MS. Maetze, A. Shannon, HS Sullivan, TJ. Norman, RW. Wells, RP. (1996). Disability resulting from occupational low back pain. Part II: What do we know about secondary prevention? A review of the scientific evidence on prevention after disability begins. Spine. 21 (24): 2918 – 2929

Fransen, M. Edmonds, J. (1999) Reliability and validity of the Euroqol in patients with osteoarthritis of the knee. Rheumatology; 38 (9): 807 – 813. https://doi:org/10.1093/rheumatology/38.9.807

Friedman, BW. Chilstrom, M. Bijur, PE. Gallagher, EJ. (2010) Diagnostic testing and treatment of low back pain in United States emergency departments: a national perspective. Spine; Nov 15; 35 (24): E1406-11. Doi: 10.1097/BRS.0b013e3181d952a5.

Fritz, JM. Childs, JD. Wainner, RS. Flynn, TW. (2012) Primary care referral of patients with low back pain to physical therapy. SPINE; 37 (25): 2114 - 2121

Fritz, JM. Irrgang, JJ. (2001) A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Physical Therapy; 81: 776 – 788

Frymoyer, JW. (1998) Back pain and sciatica. New England Journal of Medicine; 318 (5): 291 – 300. Doi: 10.1056/NEJM198802043180506

Gad El-Rab, W. Zaïane, OR. El-Hajj, M. (2017) Formalizing clinical practice guideline for clinical decision support systems. Health Informatics Journal; 23 (2): 146 -156. Doi: 10.1177/1460458216632272

Gagnier, JJ. Moher, D. Boon, H. Beyene, J. Bombardier, C. (2012) Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. BMC Medical Research Methodology; 12 (111): 1 – 15. Doi: 10.1186/1471-2288-12-111

Gatchel, RJ. Rollings, KH. (2008) Evidence informed management of chronic low back pain with cognitive behavioral therapy. Spine Journal; 8 (1): 40 – 44 Doi: 10.1016/j.spinee.2007.10.007

Gates, A. Gates, M. Duarte, G. Cary, M. Becker, M. Prediger, B. Vandermeer, B. Fernandes, RM. Pieper, D. Hartling, L. (2018) Evaluation of the reliability, usability, and applicability of AMSTAR, AMSTAR 2, and ROBIS: protocol for a descriptive study. Systematic Reviews; 7 (85)

Giles, LGF. Muller, R. (2003) Chronic spinal pain: a randomized clinical trial comparing medication, acupuncture, and spinal manipulation. Spine; 28 (14): 1490 – 1502: discussion 1502 - 1503

Giles, LGF. Muller, R. (1999) Chronic spinal pain syndromes: a clinical pilot trial comparing acupuncture, a nonsteroidal anti-inflammatory drug, and spinal manipulation. Journal of Manipulative & Physiological Therapeutics; 22 (6): 376 – 381

Glasgow, RE. Vinson, C. Chambers, D. (2012) National institutes of health approaches to dissemination and implementation science: current and future directions. American Journal of Public Health; 102 (2012): 1274 - 1281

Gleiss, A. (2017) Identifiability of components of complex interventions using factorial designs. Journal of Alternative and Complementary Medicine; 23 (8): 569 -574. Doi: 10.1089/acm.2017.0075

Glombiewski, JA. Hartwich-Tersek, J. Rief, W. (2010) Attrition in cognitive-behavioural treatment of chronic back pain. The Clinical Journal of Pain; 26 (7): 593 – 601. Doi: 10.1097/AJP.0b)13e3181e37611

Goldby, LJ. Moore, AP. Doust, J. Trew, ME. (2006) A randomized controlled trial investigating the efficiency of musculoskeletal physiotherapy on chronic back disorder. Spine; 31 (10): 1083 – 1093. Doi: 10.1097/01.brs.0000216464.37504.64

Goodchild van Hilton, L. (2015) Why it's time to publish research "failures". Elsevier: https://www.elsevier.com/connect/scientists-we-want-your-negative-results-too

Gould, D. Kelly, D. Goldstone, L. Gammon, J. (2001) Visual Analogue Scale (VAS). Journal of Clinical Nursing; 10: 697 – 706

GRADE (2013) GRADE Handbook. https://gdt.gradepro.org/app/handbook/handbook.html

GRADE (2004) Grading quality of evidence and strength of recommendations. BMJ; 328: Doi: https://doi/10.1136/bmj.328.7454.1490

Grant, J. Chittleborough, C. Taylor, AW> Dal Grande, E. Wilson, DH. Phillips, PJ. Adams, RJ. Cheek, J. Price, K. Gill, T. Ruffin, RE. North West Adelaide Health Study Team. (2006). The North West Adelaide Health Study: detailed methods and baseline segmentation of a cohort for selected chronic diseases. Epidemiologic Perspectives & Innovations; 3 (1): 4

Grimshaw, JM. Eccles, P. Lavis, JN. Hill, SJ. Squires, JE. (2012) Knowledge translation of research findings. Implementation Science; 7 (50): 1 – 17. Doi: 10.1186/1748-5908-7-50

Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR. (2005) Chapter 2. Epidemiologic Measures. Medical Epidemiology. 4th edition. New York: McGraw-Hill; 2005

Greenhalgh, T. Howick, J. Maskrey, N. (2014) Evidence based medicine: a movement in crisis. The BMJ; 348: 1 – 7. Doi: 10.1136/bmj.g3725

Greenough, C. (2016) The National back pain pathway. NHS England; https://www.england.nhs.uk/blog/charles-greenough/

Grotle, M. Garratt, AM. Krogstad, JH. Stuge, B. (2012). Reliability and construct validity of self-report questionnaires for patients with pelvic girdle pain. Physical Therapy; 92 (1): 111 – 123

Gupta, N. (2014) Ascertainment or detection bias or measurement bias can be minimized by double blinding the treatment allocation in a clinical trial. BMJ; 348: 2014; 348: g2573. Doi: 10.1136/bmj.g2573

Gupta, SK. (2011) Intention – to - treat concept: A review. Perspectives in Clinical Research; 2 (3): 109 – 112 Doi: 10.4103/2229-3485.83221

Guyatt, GH. Thorlund, K. Oxman, AD. Walter, SD. Patrick, D. Furukawa, TA. Johnston, BC. Karanicolas, P. Akl, EA. Vist, G. Kunz, R. Brozer, J.

Kupper, LL. Martin, SL. Meerpohl, JJ. Alonso-Coello, P. Christensen, R. (2013) GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. Journal of Clinical Epidemiology; 66 (2): 173 – 83. Doi: 10.1016/j.jclinepi.2012.08.001

Guyatt, GH. Oxman, AD. Sultan, S. Glasziou, P. Akl, EA. Alonso-Coello, P. Atkins, D. Kunz, R. Brozek, J. Montori, V. Jaeschke, R. Rind, D. Dahm, P. Meerpohl, J. Vist, G. Berliner, E. Norris, S. Falck-Ytter, Y. Murad, MH. Schünemann, HJ. The GRADE Working Group (2011) GRADE guidelines: 9. Rating up the quality of evidence. Journal of Clinical Epidemiology; 64 (12): 1311 – 1316. Doi: https://doi.org/10.1016/j.jclinepi.2011.06.004

Guyatt, GH. Oxman, AD. Kunz, R. Woodcock, J. Brozek, J. Helfand, M. Alonso-Coello, P. Glasziou, P. Jaeschke, R. Akl, EA. Norris, S. Vist, G. Dahm, P. Shukla, VK. Higgins, J. Falck-Ytter, Y. Schünemann, HJ. The GRADE Working Group (2011b) GRADE guidelines: 7. Rating the quality of evidence—inconsistency. Journal of Clinical Epidemiology; 64 (12): 1294 – 1302. Doi: https://doi.org/10.1016/j.jclinepi.2011.03.017

Guyatt, GH. Oxman, AD. Kunz, R. Woodcock, J. Brozek, J. Helfand, M. Alonso-Coello, P. Falck-Ytter, Y. Jaeschke, R. Vist, G. Akl, EA. Post, PN. Norris, S. Meerpohl, J. Shukla, VK. Nasser, M. Schünemann, HJ. The GRADE Working Group (2011c) GRADE guidelines: 8. Rating the quality of evidence—indirectness. Journal of Clinical Epidemiology; 64 (12): 1303 – 1310. Doi: https://doi.org/10.1016/j.jclinepi.2011.04.014

Guyatt, GH. Oxman, AD. Kunz, R. Brozek, J. Alonso-Coello, P. Rind, D. Devereaux, PJ. Montori, VM. Freyschuss, B. Vist, G. Jaeschke, R. Williams, JW. Murad, MH. Sinclair, D. Falck-Ytter, Y. Meerpohl, J. Whittington, C. Thorlund, K. Andrews, J. Schünemann, HJ. The GRADE Working Group (2011d) GRADE guidelines 6. Rating the quality of evidence—imprecision. Journal of Clinical Epidemiology; 64 (12): 1283 – 1293. Doi: https://doi.org/10.1016/j.jclinepi.2011.01.012

Haake, M. Müller, HH. Schade-Brittinger, C. Basler, HD. Schäfer, H. Maier, C. Endres, HG. Trampisch, HJ. Molsberger, A. (2007) German Acupuncture Trials (GERAC) for chronic low back pain. Archives Internal Medicine; 167 (17): 1892 – 1898

Haas, M. Vavrek, D. Peterson, D. Polissar, N. Neradilek, MB. (2014) Dose-response and efficacy of spinal manipulation for care of chronic low back pain: a randomised controlled trial. The Spine Journal; 14 (7): 1106 – 1116. Doi: 10.1016/j.spinee.2013.07.468

Hackethal, V. (2017) New ACP guidelines for nonradicular low back pain. Medscape; February 13: http://www.medscape.com/viewarticle/875737#vp_3

Hägg, O. Fritzell, P. Nordwall, A. Swedish Lumbar Spine Study Group. (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. European Spine Journal; 12: 12 – 20. Doi: 10.1007/s00586-002-0464-0

Hansen, Z. Daykin, A. Lamb, SE. (2010) A cognitive-behavioural programme for the management of low back pain in primary care: a description and justification of the intervention used in the Back Skills Training Trial (BeST; ISRCTN 54717854) Physiotherapy; 96; 87 – 94

Hartvigsen, J. Hancock, MJ. Kongsted, A. Woolf, A. et al (31 authors) (2018) What low back pain is and why we need to pay attention. The Lancet; 391 (10137) Doi: 10.1016/S0140-6736(18)30480-X

Hasegawa, TM. Baptista, AS. De Souza, MC. Yoshizumi, AM. Natour, J. (2014) Acupuncture for acute non-specific low back pain: a randomised, controlled, double-blind, placebo trial. Acupuncture in Medicine; 32 (2): 109 - 115

Hasenbring, MI. Rusa, AC. Turk, DC. (2012) From acute to chronic back pain: Risk factors, mechanisms and clinical implications. Oxford University Press inc; New York

Hawe, P. Shiell, A. Riley, T. (2004) Complex interventions: how “out of control” can a randomised controlled trial be? BMJ; 328 (7455): 1561 – 1563. Doi: 10.1136/bmj.328.7455.1561

Hawk, C. Long, CR. Reiter, R. Davis, CS. Cambron, JA. Evans, R. (2002) Issues in Planning a Placebo-Controlled Trial of Manual Methods: Results of a Pilot Study. The journal of alternative and complementary medicine; Volume 8: (1) 21 – 32

Hayes, MHS. Patterson, DG. (1921) Experimental development of the graphic rating method. Psychological Bulletin; 18: 98 – 99

He, T. Zhu, W. Du, SQ. Yang, JW. Li, F. Yang, BF, Shi, GX, Liu, CZ. (2015) Neural mechanisms of acupuncture as revealed by fMRI studies. Autonomic Neuroscience; 190: 1 – 9. Doi: 10.1016/j.autneu.2015.03.006

Health and Safety Executive (HSE) (2017) Work-related musculoskeletal disorder (WRMSDs) statistics, Great Britain. www.hse.gov.uk/statistics/index.htm

Hecker, HU. Steveling, A. Peuker, E. Kastner, J. Liebchen, K. (2001) Color atlas of acupuncture. Thieme, New York

Henderson, M. Page, L. (2007) Appraising the evidence: what is selection bias? Evidence-Based Mental Health; 10 (3): 61 Doi: 10.1136/ebmh.10.3.67

Henschke, N. Maher, CG. Ostelo, RW. de Vet, HC. Macaskill, P. Irwig, L. (2013) Red flags to screen for malignancy in patients with low-back pain. Cochrane Database of Systematic Reviews; Issue 2: CD008686. DOI: 10.1002/14651858.CD008686.pub2

Hewitt, CE. Kumaravel, B. Dumville, JC. Torgerson, DJ. (2010) Assessing the impact of attrition in randomized controlled trials. Journal of Clinical Epidemiology; Volume 63 (11) 1264 – 1270

Hickey, GL. Grant, SW. Dunning, J. Siepe, M. (2018) Statistical primer:

sample size and power calculations – why, when and how? European Journal of Cardio-Thoracic Surgery; 54 (1): 4 – 9. https://doi.org/10.1093/ejcts/ezy169

Hill, JC. Whitehurst, DGT. Lewis, M. Bryan, S. Dunn, KM. Foster, NE. Konstantinou, K. Main, CJ. Mason, E. Somerville, S. Sowden, G. Vohora, K. Hay, EM. (2011) Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. The Lancet; 378 (9802): 1560 – 1571. Doi: 10.1016/S0140-6736(11)60937-9

Hidalgo, B. Detrembleur, C. Hall, T. Mahaudens, P. Nielens, H. (2014) The efficacy of manual therapy and exercise for different stages of non-specific low back pain: an update of systematic reviews. Journal of Manipulative Therapy; 22 (2): 59 – 74. Doi: 10.1179/2042618613Y.0000000041

Higgins, JPT, Thompson, SG. (2002) Quantifying heterogeneity in a meta-analysis. Statistics in Medicine; 21 (11): 1539 - 1558. Doi: 10.1002/sim.1186

Hoffmann, TC. Glasziou, PP. Boutron, I. Milne, R. Perera, R. Moher, D. Altman, DG. Barbour, V. Macdonald, H. Johnston, M. Lamb, SE. Dixon-Woods, M. McCulloch, P. Wyatt, JC. Chan, AW. Michie, S. (2014) Better reporting of interventions: template for investigation description and replication (TIDieR) checklist and guide. BMJ; 348: g1687. Doi: 10.1136/bmj.g1687

Hoffman, BM. Papas, RK. Chatkoff, DK. Kerns, RD. (2007) Meta-analysis of psychological interventions for chronic low back pain. Health Psychology; 26 (1): 1 – 9. Doi: 10.1037/0278-6133.26.1.1

Hoiriis, KT. Pfleger, B. McDuffie, FC. Cotsonis, G. Elsangak, O. Hinson, R. Verzosa, GT. (2004) A randomized clinical trial comparing chiropractic adjustments to muscle relaxants for sub-acute low back pain. Journal of Manipulative and Physiological therapeutics; 27 (6): 388 – 398. Doi:

http://doi.org/10.1016/j.jmpt.2004.05.003

Holding, E. Relton, C. Roberts, K. Oliver, E. (2017) First aid intervention in the adult population: the Yorkshire Health Study and its implications for first aid education. International Journal of First Aid Education

Hollingworth, W. Todd, CJ. King, H. Males, T. Dixon, AR. Karia, KR. Kinmonth, AL. (2002) Primary care referrals for lumbar spine radiography: diagnostic yield and clinical guidelines. The British Journal of General Practice; June; 52 (479): 475– 80

Hopton, AK. Curnoe, S. Kanaan, M. MacPherson, H. (2012) Acupuncture in practice: mapping the providers, the patients and the settings in a national cross-sectional survey. BMJ Open; 2 (1) Doi: 10.1136/bmjopen-2011-000456

House of Lords (HoL) (2000) House of Lords, Science and technology, sixth report, 2000, www/parliament.the-stationary-office/co.uk/pa

Howard, L. Donovan, J. (2009) Why is recruitment to trials difficult? An investigation into recruitment difficulties in an RCT of supported employment in patients with severe mental illness. Contemporary Clinical Trials; 30 (1): 40 – 46. Doi: 10.1016/j.cct.2008.07.007

Hoy, D. March, L. Brooks, P. Blyth, F. Woolf, A. Bain, C. Williams, G. Smith, E. Vos, T. Barendregt, J. Murray, C. Burstein, R. Buchbinder, R. (2014) The global burden of low back pain: estimates from the global burden of disease 2010 study. Ann Rheum Dis: 73: 968 – 974 Doi: 10.1136/annrheumdis-20130204428

Hoy, D. Bains, C. Williams, G. March, L. Brooks, P. Blyth, F. Woolf, A. Vos, Theo. Buchbinder, R. (2012) A systematic review of the global prevalence of low back pain. Arthritis and Rheumatology; 64: 2028 – 2037

Hoy, D. Brooks, P. Blyth, F. Buchbinder, R. (2010) The epidemiology of low back pain. Best Practice & Research Clinical Rheumatology; 24 (6): 769 - 781

Hoy, DG. March, L. Brooks, P. Woolf, A. Blyth, F. Vos, T. Buchbinder, R. (2010) Measuring the global burden of low back pain. Best Practice & Research Clinical Rheumatology; 24: 155–65

Hróbjartsson, A. Thomsen, ASS. Emanuelsson, F. Tendal, B. Hilden, Hildon, J. Boutron, I. Ravaud, P. Brorson, S. (2012) Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ; 344: 1 – 11. Doi: 10.1136/bmj.e1119

Huang, GD. Bull, J. McKee, KJ. Mahon, E. Harper, B. Roberts, JN. Clinical trials recruitment planning: A proposed framework from Clinical Trials Transformation Initiative. Contemporary Clinical Trials; 66: 74 – 79

Huo, T. Guo, Y. Shenkman, E. Muller, K. (2018) Assessing the reliability of the short form 12 (SF-12) health survey in adults with mental health conditions: a report from the wellness incentive and navigation (WIN) study. Health and Quality of Life Outcomes; 16: 34.Doi: 10.1186/s12955-018-0858-2

Hutchinson, AJP. Ball, S. Andrews, JCH. Jones, GG. (2012) The effectiveness of acupuncture in treating chronic non-specific low back pain: a systematic review of the literature. Journal of Orthopedic Surgery and Research; 7: 36 Doi: 10.1186/1749-799X-7-36

Hurwitz, EL. Morgenstern, H. Kominski, GF. Yu, F. Chaing, LM. (2006) A randomized trial of chiropractic and medical care for patients with low back pain: eighteen-month follow-up outcomes from the UCLA low back pain study. Spine; 31 (6): 611 – 621. Doi: 10.1097/01.brs.0000202559.41193.b2

Inoue, M. Kitakoji, H. Ishizaki, N. Tawa, M. Yano, T. Katsumi, Y. (2006) Relief of low back pain immediately after acupuncture treatment- a randomised, placebo controlled trial. Acupuncture in Medicine; 24 (3): 103 - 108

Inoue, M. Hojo, T. Yano, T. Katsumi, Y. (2005) Effects of lumbar acupuncture stimulation on blood flow to the sciatic nerve – an exploratory study. Acupuncture in Medicine; 23 (4): 166 – 170. Doi: 10.1136/aim.23.4.166

Institute for Clinical Systems Improvement (ICSI) (2005) Adult Low Back Pain. Bloomington, MN: ICSI

James ND, Sydes MR, Clarke NW, Mason MD, Dearnaley DP, Anderson J, Popert, R. Sanders, K. Morgan, R. Stansfeld, J. Dwyer, J. Masters, J. Parmer, MKB. (2008) STAMPEDE: Systemic Therapy for Advancing or Metastatic Prostate Cancer – A Multi-Arm Multi-Stage Randomised Controlled Trial. Clinical Oncology; 20 (8): 577 – 581

Janssen, M. Pickard, A. Golicki, D. Gudex, C. Niewada, M. Scalone, L. Swinburn, P. Busschbach, J. (2012). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Quality of Life Research; 22 (7): 1717 – 1727

Jaques, E. (2018) Idiopathic pain is well known even if its cause isn't. Very well health, 18th May 2018, medically reviewed by a board-certified physician. www.verywellhealth.com/what-is-idiopathic-pain-2564613

Jensen, MP. Karoly, P. (2011) "Self-report scales and procedures for assessing pain in adults" in Turk, D.C. and Melzack, R. (eds) Handbook of pain assessment, 3rd edition. New York, Guilford Press, 19 – 44

Jin, H. Guo, J. Liu, J. Lyu. B. Foreman, RD. Yin, J. Shi, Z. Chen, JDZ. (2017) Anti-inflammatory effects and mechanism of vagal nerve stimulation combined with electro-acupuncture in a rodent model of TNBS-induced colitis. American Journal of Physiology, Gastrointestinal and Liver Physiology; 313 (3): G192 – G202. Doi: 10.1152/ajpgi.00254.2016

Jin, L. Sun, J. Xu, Z. Yang, X. Liu, Y. Qin, W. (2016) Intersubject synchronization analysis of brain activity associated with the instant

effects of acupuncture: an fMRI study. Acupuncture in Medicine. 36 (1): 14 – 20. Doi: 10.1136/acupmed-2016-011327

Johnson, J. Coons, S. (2017). Comparison of the EQ-5D and SF-12 in an adult. Quality of Life Research; 7: 155 – 166

Johnsen, LG. Hellum, C. Nygaard, OP. Storheim, K. Brox, JI. Rossvoll, I. Leivseth, G. Grotle, M. (2013) Comparison of the SF6D, the EQ5D, and the Oswestry Disability Index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskeletal Disorders; 14: 148. Doi: 10.1186/1471-2474-14-148

Joos, E. Peretz, A. Beguin, S. Famaey, JP. (1991) Reliability and reproducibility of visual analogue scale and numeric rating scale for therapeutic evaluation of pain in rheumatic patients, Journal of Rheumatology; 18 (8): 1269 – 1270

Jordan, K. Dunn, K. M. Lewis, M. Croft, P. (2006) A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. Journal of Clinical Epidemiology; 59 (1): 45 – 52

Juch, JNS. Maas, ET. Ostelo, RWJG. Groeneweg, JG. Kallewaard, JW. Koes, BW. Verhagen, AP. Van Dongen, JM. Huygen, FJPM. Van Tulder, MW. (2017) Effect of radiofrequency denervation on pain intensity among patients with chronic low back pain. JAMA; 318 (1): 68 – 81. Doi: 10.1001/jama.2017.7918

Julious, SA. (2005). Sample size of 12 per group, rule of thumb for a pilot study. Pharmaceutical Statistics; 4: 287 – 291

Juniper, M. Le, TK. Mladsi, D. (2009) The epidemiology, economic burden and pharmacological treatment of chronic low back pain in France, Germany, Italy, Spain and the UK: a literature based review. Expert Opinion on Pharmacotherapy; 10 (16): 2581 -2592. Doi: 10.1517/14656560903304063

Jütte, R. (2013) The early history of placebo. Complementary Therapies in Medicine; 21 (2): 94 – 97. Doi: 10.1016/j.ctim.2012.06.002

Kamper, SJ. Williams, CM. (2013) The placebo effect: powerful, powerless or redundant. British Journal of Sports Medicine; 47 (1): 6 – 9 Doi: http://dx.doi.org/10.1136/bjsports-2012-091472

Kang, M. Ragan, BG. Jae-Hyeon, P. (2008) Issues in outcome research: An overview of randomization techniques for clinical trials. Journal of Athletic Training; 43 (2): 215 – 221

Kaptchuk, TJ. Miller, FG. (2018) Open label placebo: can honesty prescribed placebos evoke meaningful therapeutic benefits? BMJ; 363. Doi: https://doi.org/10.1136/bmj.k3889

Kaptchuk, TJ. Kelley, JM. Conboy, LA. Davis, RB. Kerr, CE. Jacobson, EE. Kirsch, I. Schyner, RN. Nam, BM. Nguyen, LT. Park, M. Rivers, AL. McManus, C. Kokkotou, E. Drossman, DA. Goldman, P. Lembo, AJ. (2008) Components of the placebo effect: a randomized controlled trial in irritable bowel syndrome. BMJ; 336 (998): 1003. Doi: 10.1136/bmj.39524.439618.25

Kaptchuk, TJ. (2000) The double-blind, randomized, placebo-controlled trial:Gold standard or golden calf? Journal of Clinical Epidemiology; 54 (6): 541 – 549

Katz, NP. Paillard, FC. Ekman, E. (2015) Determining the clinical importance of treatment benefits for interventions for painful orthopaedic conditions. Journal of Orthopaedic Surgery and Research; 10: 24. Doi: 10.1186/s13018-014-0144

Kennedy, S. Baxter, GD. Kerr, DP. Bradbury, I. Park, J. McDonough, SM. (2008) Acupuncture for acute non-specific low back pain: a pilot randomised non-penetrating sham controlled trial. Complementary Therapies in Medicine; 16 (3): 139 - 146

Kennedy-Martin, T. Curtis, S. Faries, D. Robinson, S. Johnston, J. (2015)

A literature review on the representativeness of randomized controlled trials samples and implications for external validity of trial results. Trials; 16: 495. Doi: 10.1186/s13063-015-1023-4

Kennelly, B. (2017) How should cost-of-illness studies be interpreted? The Lancet; Psychiatry; 4 (10): 735 – 736 Doi: 10.1016/S2215-0366(17)30364-4

Keränen, T. Halkoaho, A. Itkonen, E. Pietillä, AM. (2015) Placebo-controlled clinical trials: how trial documents justify the use of randomisation and placebo. BMC Medical Ethics; 16: 2 Doi: 10.1186/1472-6939-16-2

Kim, SYH. Flory, J. Relton, C. (2018) Ethics and practice of Trials within Cohorts: An emerging pragmatic trial design. Clinical Trials; 15 (1): 9 – 16. Doi: 10.1177/1740774517746620

Kim, SJ. Min, S. Lee, H. Cheon, S. Zhang, X. Park, JY. Song, TJ. Park, HJ. (2016) Changes of local blood flow in response to acupuncture stimulation: A systematic review. Evidence-Based Complementary and Alternative Medicine; 9874207: 1 - 11. Doi: 10.1155/2016/9874207

Kim, KH. Kim, TH. Lee, BR. Kim, JK. Son, DW. Lee, SW. Yang, GY. (2013) Acupuncture for lumbar spinal stenosis: a systematic review and meta-analysis. Complementary Therapies in Medicine; 21 (5): 535 - 556

Kim, KH. Lee, H. Chae, Y. Park, HJ. Lee, H. (2012) A systematic review of cost-effectiveness analysis alongside randomised controlled trials of acupuncture. Acupuncture in Medicine 30 (4), 273-285

KJoller, M. Rasmussen, NK. Keilding L. Petersen, HC. Nielsen, GA. (1995), Sundhed og sygelighed i Danmark 1994 -og udviklingen siden 1987. Dansk Inst Klinisk Epidemiology. Cited in Dufour, N. Thamsborg, G. Oefeldt, A. Lundsgaard, C. Stender, S. (2010) Treatment of Chronic Low Back Pain: A Randomized, Clinical Trial Comparing Group-Based Multidisciplinary Biopsychosocial Rehabilitation and Intensive Individual

Therapist-Assisted Back Muscle Strengthening Exercises. Spine; 35 (5): 469 - 476

Kjøller M, Rasmussen NK, (1995): Health and morbidity in Denmark 2000 and the development since 1987. Copenhagen 2002: National Institute of Public Health, Øster Farimagsgade 5A, 2, 1399 Copenhagen K, Denmark

Klaber-Moffet, J. Jackson, DA. Gardiner, ED. Torgerson, DJ. Coulton, S. Eaton, S. Mooney, MP. Pickering, C. Green, AJ. Walker, LG. May, S. Young, S. (2006) Randomized trial of two physiotherapy interventions for primary care neck and back pain patients: 'McKenzie' V's brief physiotherapy pain management Rheumatology; 45: 1514 – 1521 Doi: 10.1093/rheumatology/kel339

Klaber-Moffet, J.A. Jackson, D.A. Richmond, S. Hahn, S. Coulton, S. Farrin, A. Manca, A. Torgerson, D. (2005) Randomised trial of a brief physiotherapy intervention compared with usual physiotherapy for neck pain patients: outcomes and patients' preference. British Medical Journal (BMJ); 330, 75 - 78

Klaber-Moffet, J. Torgerson, D. Bell-Syer, S. Jackson, D. Llewlyn-Phillips, H. Farrin, A. Barber, J. (1999) Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. British Medical Journal; 319: 279 – 283

KNGF (Koninklijk Nederlands Genootschap voor Fysiotherapie) (2013) Clinical practice guideline for physical therapy in low back pain. Practice Guidelines. Royal Dutch Society of Physical Therapy. V-07/2013. http://www.ipts.org.il/_Uploads/dbsAttachedFiles/low_back_pain_practice_guidelines_2013.pdf

Koes, BW. Van Tulder, M. Lin, CWC. Macedo, LG. McAuley, J. Maher C. (2010) An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. European Spine Journal; 19: 2075-2094

Koes, BW. Van Tulder, MW. Thomas, S. (2006) Diagnosis and treatment of low back pain. BMJ; 332 Doi: https://doi.org/10.1136/bmj.332.7555.1430

Koes, BW. Van Tulder, MW. Ostelo, R. Burton, K. Waddell, G. (2001) Clinical guidelines for the management of low back pain in primary care: an international comparison. Spine; 26 (22): 2504 – 2513

Koes BW, Bouter LM, van Mameren H, Essers AH, Verstegen GM, Hofhuizen DM, Houben JP, Knipschild PG. (1992) The effectiveness of manual therapy, physiotherapy, and treatment by the general practitioner for nonspecific back and neck complaints. A randomized clinical trial. Spine; 17 (1): 28 – 35

Koletsi, D. Pandis, N. Polychronopoulou, A. Eliades, T. (2012) What's in a title? An assessment of whether randomized controlled trial in a title means that it is one. American Journal of Orthodontics and Dentofacial Orthopedics; 141 (6): 679 – 685. Doi: 10.1016/j.ajodo.2011.12.020

Konstantinou, K. Dunn, KM. (2008) Sciatica: Review of epidemiological studies and prevalence estimates. Spine; 33 (22): 2464 – 2472. Doi: 10.1097/BRS.0b013e318183a4a2

Kovacs, FM. Abraira, V. Zamora, J. Teresa Gil del Real, M. Llobera, J. Fernández, C. Bauza, JR. Bauza, K. Coll, J. Cuadri, M. Duro, E. Gili, J. Gestoso, M. González, J. Ibañez, P. Jover, A. Lázaro, P. Llinás, M. Mateu, C. Mufraggi, N. Muriel, A. Nicolau, C. Olivera, MA. Pascual, P. Perelló, L. Pozo, F. Revuelta, T. Reyes, V. Ribot, S. Ripoll, J. Ripoll, J. Rodríguez, E. Kovacs-Atención Primaria Group. (2004) Correlation between pain, disability, and quality of life in patients with common low back pain. Spine; 29 (2): 206 – 201. Doi: 10.1097/01.BRS.0000107235.47465.08

Kuczynski, JJ. Schwieterman, B. Columber, K. Knupp, D. Shaub, L. Cook, CE. (2012) Effectiveness of physical therapist administered spinal manipulation for the treatment of low back pain: a systematic review of the literature. International Journal Sports Physical Therapy; Dec 7 (6):

647 – 662

Kühne, F. Ehmcke, R. Härter, M. Kriston, L. (2015) Conceptual decomposition of complex health care interventions for evidence synthesis: a literature review. Journal of Evaluation of Clinical Practice; 21 (5): 817 – 823. Doi: 10.1111/jep.12384

Kunz, R. Oxman, A, D. (1998) The unpredictability paradox: review of empirical comparisons of randomized and non-randomised clinical trials. British Medical Journal (BMJ); 317: 1185 - 1190

Kumar, S. Beaton, K. Hughes, T. (2013) The effectiveness of massage therapy for the treatment of nonspecific low back pain: a systematic review of systematic reviews. International Journal of General Medicine; 2013; 6: 733–741. Sep 4. Doi: 10.2147/IJGM.S50243

Kumnerddee, W. (2009) Effectiveness comparison between Thai traditional massage and Chinese acupuncture for myofascial back pain in Thai military personnel: a preliminary report. Journal of the Medical Association of Thailand; 92 (1): 117

Kyte, DG. Calvert, M. van der Wees, PJ. Hove, R. Tolan, S. Hill, JC. (2015) An introduction to patient-reported outcome measures (PROMs) in physiotherapy. Physiotherapy; 10: (2): June: 119–125

Kwon, Y. Lee, S. Lee, C. Jung, S. Kim, D. Choi, S. (2007) The short-term efficacy of acupuncture for chronic low back pain: randomised sham controlled trial. Journal of Oriental Rehabilitation Medicine; 17 (2): 123 - 132

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

Lateef, H. Patel, D. (2009) What is the role of imaging in acute low back pain. Current reviews in musculoskeletal medicine; 2 (2): 69 – 73 Doi: 10.1007/s12178-008-9037-0

Lam, M. Galvin, R. Curry, P. (2013) Effectiveness of acupuncture for nonspecific chronic low back pain: a systematic review and meta-analysis. Spine; 38 (24): 2124 - 2138

Lamb. SE. Hansen, Z. Lall, R. Castelnuovo, E. Withers, EJ. Nichols, V. (2010) Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. Lancet; 375: 916e23

Lancaster, GA. Dodd, S. Williamson, PR. (2004) Design and analysis of pilot studies: recommendations of good practice. Journal of Evaluation in Clinical Practice; 10: 307 -312

Langevin, HM. (2014) Acupuncture, connective tissue, and peripheral sensory modulation. Critical Reviews in Eukaryotic gene expression; 24 (3): 249 – 253.

Langevin, HM. Bouffard, NA. Churchill DL. Badger, GL. (2007) Connective tissue fibroblast response to acupuncture: dose dependant effect of bidirectional needle rotation. Journal of Alternative & Complimentary Medicine 2007. 13 (3): 355

Langevin, HM. Hammerschlag, R. Lao, L. Napadow, V. Schnyer, RN. Sherman, KJ. (2006) Controversies in Acupuncture Research: Selection of Controls and Outcome Measures in Acupuncture Clinical Trials. The Journal of Alternative and Complementary Medicine; 12 (10): 943 – 953

Langevin, HM. Yandow, JA. (2002) Relationship of acupuncture points and meridians to connective tissue planes. The Anatomical Record; 269 (6): 257 – 265. Doi: 10.1002/ar.10185

Lansberg, MG. Bhat, NS. Yeatts, SD. Palesch, YY. Broderick, JP. Albers, GW. Lai, TL. Lavori, PW. (2016) Power of adaptive trial design for endovascular stroke studies. Stroke; 47: 2931 – 2937 Doi: 10.1161/STROKEAHA.116.015436

Lauridsen, HH. Hartvigsen, J. Manniche, C. Korsholm. L. Grunnet-Nilsson, N. (2006) Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. BMC Musculoskeletal Disorders; 7 (82)

Lee, JH. Choi, TY. Lee, MS. Lee, H. Shin, BC. Lee, H. (2013) Acupuncture for Low back pain: a systematic review. Clinical Journal of Pain; 29 (2): 172 – 185

Lee, C. Crawford, C. Wallerstedt, D. York, A. Dincan, A. Smith, J. Sprengel, M. Welton, R. Jonas, W. (2012) The effectiveness of acupuncture research across components of the trauma spectrum response (tsr): a systematic review of reviews. Systematic Reviews; 1 (46): 1 – 18. Doi: 10.1186/2046-4053-1-46

Lehtinen, P. FICEBO Group. 2018 (2018) Most common shoulder operation is no more beneficial than placebo surgery. Science Daily: 20th July 2018; University of Helsinki www.sciencedaily.com/releases/2018/07/180720092518.htm

Leibing, E. Leonhardt, U. Köster, G. Ramadori, G. (2002) Acupuncture treatment of chronic low-back pain - a randomized, blinded, placebo-controlled trial with 9 -month follow up. Pain; 96 (1 - 2): 189 - 196 Doi: 10.1016/S03043959(01)00444-4

Leon, AC. Mallinckrodt, CH. Chuang-Stein, C. Archibald, DG. Archer, GE. Chartier, K. (2006) Attrition in Randomized Controlled Clinical Trials: Methodological Issues in Psychopharmacology. Biological Psychiatry; 59: 1001–1005

Lestini, G. Dumont, C. Mentré, F. (2015) Influence of the size of cohorts in adaptive design for nonlinear mixed effects models: An evaluation by simulation for a pharmacokinetic and pharmacodynamics model for a biomarker in oncology. Pharmaceutical Research; 32 (10): 3159 – 3169 Doi: 10.1007/s11095-015-1693-3

Lewis, K. Abdi, S. (2010) Acupuncture for LBP: A review. Clinical Journal of Pain: Jan 26: (1): 60 – 69

Lewith, GT. White, JP. Pariente, J. (2005) Investigating acupuncture using brain imaging techniques: The current state of play. Evidence-Based Complementary and Alternative Medicine; 2 (3): 315 – 319. Doi: 10.1093/ecam/neh110

Liang, YD. Li, Y. Zhao, J. Wang, XY. Zhu, HZ. Chen, XH. (2017) Study of acupuncture for low back pain in recent 20 years: a bibliometric analysis via CiteSpace. Journal of Pain Research; 10: 951 – 964. Doi: 10.2147/JPR.S132808

Li, X. Wang, R. Shi, X. Chen, Z. Pan, Y. Li, X. Yang, K. (2016) Reporting characteristics and risk of bias in randomised controlled trials of acupuncture analgesia published in PubMed-listed journals. Acupuncture in Medicine; 35 (4): Doi: 10.1136/acupmed-2016-011149

Li, QQ. Shi, GX. Xu, Q. Wang, J. Lui, CZ. Wang, LP. (2013) Acupuncture effect and central autonomic regulation. Evidence-Based Complementary and Alternative Medicine; 2013: 267959. Doi: 10.1155/2013/267959

Liberati, A. Altman, DG. Tetzlaff, J. Mulrow, C. Gøtzsche, PC. Ioannidis, JPA. Clark, M. Devereaux, PJ. Kleijnen, J. Moher, D. (2009). "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration." BMJ; 339: (jul21_1): b2700

Lidgren, L. (2003) The bone and joint decade 2000–2010. Bulletin of the World Health Organisation; 81: 629

Lilienfield, AM. (1982) Ceterius parabus: the evolution of the clinical trial. Bulletin of the history of medicine; 56: 1 – 18

Lim, HD. Kim, MH. Lee, CY. (2016) Anti-inflammatory effects of acupuncture stimulation via the vagus nerve. Journal PLOS one; March 18: Doi: 10.1371/journal.pone.0151882

Lin, JG. Chen, CH. Huang, YC. Chen, YH. (2012) How to design the control group in randomized controlled trials of acupuncture? Evidence-Based Complementary and Alternative Medicine; 875284. Doi: 10.1155/2012/875284

Linde, K. Niemann, K. Meissner, K. (2010) Are sham acupuncture interventions more effective than (other) placebos? A re-analysis of data from the Cochrane review on placebo effects. Forsch Komplementmed; 17 (5): 259 -264

Liu, L. Skinner, M. McDonough, S. Mabire, L. Baxter, GD. (2015) Acupuncture for Low Back Pain: An Overview of Systematic Reviews. Evidence-Based Complementary and Alternative Medicine; ID 328196: 18. http://dx.doi.org/10.1155/2015/328196

Little, P. Lewith, G. Webley, F. Evans, M. Beattie, A. Middleton, K. Barnett, J. Ballard, K. Oxford, F. Smith, P. Yardley, L. Hollinghurst, S. Sharp, D. (2008) Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain. BMJ; 337: Doi: https://doi.org/10.1136/bmj.a884

Livingston, C. King, V. Little, Pettinari, C. Thielke, A. Gordon, C. (2011) Evidence-based clinical guidelines project. Evaluation and management of low back pain: A clinical practice guideline based on the joint practice guideline of the American College of Physicians and the American Pain Society (Salem, Oregon: Office for Oregon Health Policy and Research)

Loannidis, JP. Adami, HO. (2008) Nested randomized trials in large cohorts and biobanks: studying the health effects of lifestyle factors. Epidemiology; 19 (1): 75 – 82. Doi: 10.1097/EDE.0b013e31815be01c

Longo, UG. Loppini, M. Denaro, L. Maffulli, N. Denaro, V. (2010) Rating scales for low back pain. British medical bulletin; 94 (1): 81 – 144. Doi: 10.1093/bmb/ldp052

Lonnemann, E. Widerman, T. Kamper, S. Cook, C. (2012) How do I

evaluate and apply the numerous guidelines for low back pain; a practical and informed approach for clinicians. IFOMPT; Quebec City, Canada

Lowe, C. Aiken, A. Day, AG. Depew, W. Vanner, SJ. (2017) Sham acupuncture is as efficacious as true acupuncture for the treatment of IBS: A randomized placebo controlled trial. Neurogastroenterology and Motility; 29 (7) Doi: 10.1111/nmo.13040

Lozano, R. Naghavi, M. Foreman, K. Lim, S. Shibuya, K. Aboyans, V. Abraham, J. Adair, T. Aggarwal, R. Ahn, SY. Alvardo, M. Anderson, HR. et al. (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 15; 380 (9859): 2095-128. Doi: 10.1016/S0140-6736(12)61728-0

Lund, I. Näslund, J. Lundeberg, T (2009) Minimal acupuncture is not a valid placebo control in randomised controlled trials of acupuncture: a physiologist's perspective. Chinese Medicine; 4:1

Lund, I. Lundeberg, T. (2006) Are minimal, superficial or sham acupuncture procedures acceptable as inert placebo controls? Education, practice and debate. Acupuncture in Medicine; 24:13-15 Doi: 10.1136/aim.24.1.13

Lundeberg, T. Lund, I. Sing, A. Näslund, J. (2011) Is Placebo Acupuncture What It Is Intended to Be? Evidence-Based Complementary and Alternative Medicine: 932407 Doi: 10.1093/ecam/nep049

Luo, N. Chew, LH. Fong, KY. Koh, DR. Ng, SC. Yoon, KH. Vasoo, S. Li, SC. Thumboo, J. (2003) Validity and reliability of the Eq-5D self-report questionnaire in English-speaking Asian patients with rheumatic diseases in Singapore. Quality of Life Research; 12: 87 - 92

MacPherson, H. (2017a) NICE for some interventions, but not so nice for others: Questionable guidance on acupuncture for osteoarthritis and low-

back pain – invited commentary. The Journal of Alternative and Complementary Medicine; 23 (4): 247 – 248. Doi 10.1089/acm.2017.0029

MacPherson, H. Vertosick, EA. Foster, NE. Lewith, G. Linde, K. Sherman, KJ. Witt, CM. Vickers, AJ. (2017b) The persistence of the effects of acupuncture after a course of treatment: a meta-analysis of patients with chronic pain. PAIN; The Journal of the International Association for the Study of Pain; 158 (5): 784 – 793. Doi: 10.1097/j.pain0000000000000747

MacPherson, H. Vickers, AJ. Bland, M. Torgerson, D. Corbett, M. Spackman, E. Saramago, P. Woods, B. Weatherly, P. Sculpher, M. Manca, A. Richmond, S. Hopton, A. Eldred, J. Watt, I. (2017c) Acupuncture for chronic pain and depression in primary care: a programme of research. Southampton (UK): NIHR Journals Library; (Programme Grants for Applied Research) Chapter 2, Acupuncture for chronic pain: an individual patient data meta-analysis. www.ncbi.nlm.nih.gov/books/NBK409498/

MacPherson, H. Vertosick, E. Lewith, G. Linde, K. Sherman, KJ. Witt, CM. Vickers, AJ. (2014) Influence of control group on effect size in trials of acupuncture for chronic pain: A secondary analysis of an individual patient data meta-analysis. PLoS One; 9 (4): e93739 Doi: 10.1371/journal.pone.0093739

MacPherson, H. Maschino, AC. Lewith, G. Foster, NE. Witt, CM. Vickers, AJ. (2013) Characteristics of acupuncture treatment associated with outcome: An individual patient meta-analysis of 17,922 patients with chronic pain in randomised controlled trials. PLoS One; 8 (10): e77438. Doi: 10.1371/journal.pone.0077438

MacPherson, H. Altman, DG. Hammerschlag, R. Youping, L. Taixiang, W. White, A. Moher, D. STRICTA Revision Group. (2010) Revised Standards of reporting for interventions in clinical trials of acupuncture (STRICTA). Journal of Evidence Based medicine. August; 3 (3): 140 -155

MacPherson, H. Thomas, K. Armstrong, B. de Valois, B. Relton, C. Mullinger, B. White, A. Flower, A. Scheid, V. (2008) Developing research strategies in complementary and alternative medicine. Complementary Therapies in Medicine; 16: 359 – 362

MacPherson, H. Thomas, K. (2007) Traditional acupuncture for low back pain: developing high quality evidence while maintaining the integrity of the intervention. European Journal of Oriental Medicine; 5 (4): 26 – 30

MacPherson, H., Scullion, A., Thomas, K. and Walters, S. (2004) Patient reports of adverse events associated with acupuncture treatment: a prospective national survey. Quality and Safety in Health Care; 13: 349 - 355

MacPherson, H. (2004). Pragmatic clinical trials. Complementary Therapies in Medicine, 12, 136-140

MacPherson, H. White, A. Cummings, M. Jobst, KA. Rose, K. Niemtzow, RC. STRICTA Group. (2002) Standards for reporting interventions in controlled trials of acupuncture: the STRICTA recommendations. Journal of Alternative and Complementary Medicine; 8 (1): 85 – 89. Doi: 10.1089/107555302753507212

MacPherson, H. Thomas, K. Walters, S. Fitter, M. (2001) "The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists." BMJ 323 (7311): 486 – 487 Doi: 10.1136/bmj.323.7311.486

MacPherson, H. White, A. Cummings, M. Jobst, K. Rose, K. Niemtzow, R. (2001) Standards for reporting interventions in controlled trials of acupuncture: the STRICTA recommendations. Complementary Therapies in Medicine; 9: 246 – 249

Maeda, Y. Kim, H. Ketter, N. Kim, J. Cina, S. Malatesta, C. Gerber, J. McManus, C. Ong-Sutherland, R. Mezzacappa, P. Libby, A. Mawla, I. Morse, LR. Kaptchuk, TJ. Audette, J. Napodow, V. (2017) Rewiring the

primary somatosensory cortex in carpal tunnel syndrome with acupuncture. Brain; 140 (4): 914 – 927. Doi: 10.1093/brain/awx015

Main, CJ. George, SZ. (2011) Psychological influences on low back pain: why you should care? Physical Therapy; 91 (5): 609 - 613 https://doi.org/10.2522/ptj.2011.91.5.609

Mahajan, R. Gupta, K. (2010) Adaptive design clinical trials: Methodology, challenges and prospect. Indian Journal of Pharmacology; 42 (4): 201 – 207 Doi: 10.4103/0253-7613.68417

Makary, MM. Lee, J. Lee, E. Eun, S. Kim, J. Jahng, GH. Kim, K. Youn, YS. Lee, JH. Park, K. (2018) Phantom acupuncture induces placebo credibility and vicarious sensations; A parallel fMRI study of low back pain patients. Scientific Reports; 8 (930). https://www.nature.com/articles/s41598-017-18870-1.pdf

Malmivaara, A. Erkintalo, M. Jousimaa, J. Kumpulainen, T. Kuukkanen, T. Pohjolainen, T. Seitsalo, S. Österman, H. (2008) Aikuisten alaselkäsairaudet. (Low back pain among adults. An update within the Finnish Current Care guidelines. Working group by the Finnish Medical Society Duodecim and the Societas Medicinae Physicalis et Rehabilitationis, Fenniae. Duodecim 124:2237–2239

Manchikanti, L. Singh, V. Falco, FJE. Benyamin, RM. Hisch, JA. (2014) Epidemiology of Low Back Pain in Adults. Neuromodulation: Technology at the Neural Interface, 17: 3–10. Doi: 10.1111/ner.12018

Maniadakis, N. Gray, A. (2000) The economic burden of back pain in the UK. Pain; 84: 95-103

Mahajan, R. Gupta1, K. (2010) Adaptive design clinical trials: Methodology, challenges and prospect. Indian Journal of Pharmacology; 42(4): 201–207. Doi: 10.4103/0253-7613.68417 PMCID: PMC2941608

Manheirmer, E. (2012) Summary of Findings Tables: Presenting the Main Findings of Cochrane Complementary and Alternative Medicine–related

Reviews in a Transparent and Simple Tabular Format. Global Advances in Health and Medicine: Improving Healthcare Outcomes Worldwide; 1(1): 90 – 1. Doi: 10.7453/gahmj.2012.1.1.015

Man-Son-Hing, M. Laupacis, A. O'Rouke, K. Molnar, FJ. Mahon, J. Chan, KB. Wells, G. (2002) Determining the clinical importance of study results. Journal of General Internal Medicine; 17 (6): 469 -476

Marignan, M. (2014) Auriculotherapy treatment protocol for low-back pain: A randomized control trial. Medical Acupuncture; 26 (3): 154 - 160

Martin, BI. Deyo, RA. Mirza, SK. Turner, JA. Comstock, BA. Hollingworth, W. Sullervan, SD. (2008) Expenditures and health status among adults with back and neck problems. JAMA; 299: 656 – 664

Maughan, EF. Lewis, JS. (2010) Outcome measures in chronic low back pain. European Spine Journal; 19 (9): 1484 – 1494. Doi: 10.1007/s00586-010-1353-6

McAlistair, FA. Straus, SE. Sackett, DL. Altman, DG. (2003) Analysis and reporting of factorial trials: a systematic review. Journal of American Medical Association; 289: 2545 – 2553

McCambridge, J. Sorhaindo, A. Quirk, A. Nanchahal, K. (2014) Patient preferences and performance bias in a weight loss trial with a usual care arm. Patient education and Counselling; 95 (2): 243 – 247. Doi: 10.1016/j.pec2014.01.003

McCormick, A, Fleming, D. Charlton, J. (1995) Morbidity statistics from general practice. Fourth national study 1991 – 1992. London: HMSO

McDonald, AM. Knight, RC. Campbell, MK. Entwistle, VA. Grant, AM. Cook, JA. Elbourne, DR. Francis, D. Garcia, J. Roberts, I. Snowden, C. (2006) What influences recruitment to randomized controlled trials? A review of trials funded by two UK funding agencies. Trials: 7: 9 Doi: 10.1186/1745-6215-7-9

McIntosh, G. Hall, H. (2011) Low back Pain (acute) BMJ Clinical Evidence, 2011, May 9, 1102

Mercer, C. Jackson, A. Hettinga, D. et al. (2006) Clinical guidelines for the physiotherapy management of persistent low back pain (LBP). London, UK: Chartered Society of Physiotherapy

MRC – Medical Research Council (MRC) (2018) Causal models of complex interventions. https://www.mrc.ac.uk/funding/how-we-fund-research/highlight-notices/causal-models-of-complex-interventions

MRC – Medical Research Council (MRC) Craig, P. Dieppe, P. Macintyre, S. Michie, S. Nazareth, I. Petticrew, M. (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ; 337

MRC/BHF - Medical Research Council (MRC) /BHF (British Heart Foundation) (2002) Heart protection study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomized placebo-controlled trial. Lancet 360: 23 – 33Meucci, RD. Fassa, AG. Faria, NMX. (2015) Prevalence of chronic low back pain: systematic review. Revista de Saúde Pública; 49: 1

Miekisiak, G. Kollataj, M. Dobrogowski, J. Kloc, W. Libionka, W. Banach, M. Lakta, D. Sobolewski, T. Sulewski, A. Nowakowski, A. Kiwic, G. Pala, A. Potaczek, T. Gierlotka, M. (2013). Validation and cross-cultural adaptation of the Polish version of the Oswestry Disability Index. Spine (Phila Pa 1976); 38 (4): E237 – 243

Milczarek, M. (2009) Literature Review: Acupuncture and the Treatment of Low Back Pain: An Evidence-Based Literature Review. Journal of the Acupuncture Association of Chartered Physiotherapists. Autumn, 39-44

Mills, EJ. Chan, A. Wu, P. Vali, A. Guyatt, GH. Altman, DG. (2009) Design, analysis and presentation of crossover trials. Trials; 10 (27): 1 – 6. Doi: 10.1186/1745-6215-10-27

Misra, DP. Agarwal, V. (2018) Systematic reviews: Challenges for their justification, related comprehensive searches and implications. Journal of Korean Medical Science; 33 (12): Doi: 10.3346/jkms.2018.33.e92

Mlinaric, A. Horvat, M. Smolčić, VŠ. (2017) Dealing with the positive publication bias: Why you should really publish your negative results. Biochemia Medica; 27 (3). Doi: 10.11613/BM.2017.030201

Moffet, HH. (2009) Sham acupuncture can be as efficacious as true acupuncture: a systematic review of clinical trials. Journal of Alternative and Complimentary Medicine; 15 (3): 213 – 216. Doi: 10.1089/acm.2008.0356

Moher, D. Hopewell, S. Schulz, KF. Montori, V. Gøtzsche, PC. Devereaux, PJ. Elbourne, D. Egger, M. Altman, DG. (2010) CONSORT 2010 Explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ; 340: 869. Doi: https://doi.org/10.1136/bmj.c869

Moher, D. Schulz, KF. Altman, DG. (2001) The CONSORT statement: revised recommendations for improving quality of reports of parallel-group randomised trials. The Lancet. 357 (9263): 1191-1194

Moher, D. Pham, B. Jones, A. (1998) Does the quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analysis? Lancet, 354: 1896-1900

Molsberger, A. Mau, J. Pawelec, DB. Winkler, J. (2002) Does acupuncture improve the orthopaedic management of chronic low back pain: A randomised blinded controlled trial with 3 months follow up. Pain; 99 (3): 579 – 587

Montgomery, AA. Peters, TJ. Little, P. (2003) Design, analysis and presentation of factorial randomised controlled trials. BMC Medical Research Methodology; 3 (26): 1 – 5. Doi: 10.1186/1471-2288-3-26

Montori, VM. Devereaux, PJ. Adhikari, NK. Burns, KE. Eggert, CH.

Briel, M. Lacchetti, C. Leung, TW. Darling, E. Bryant, DM. Bucher, HC. Schünemann, HJ. Meade, MO. Cook, DJ. Erwin, PJ. Sood, A. Sood, R. Lo, B. Thompson, CA. Zhou, Q. Mills, E. Guyatt, GH. (2005) Randomized trials stopped early for benefit: a systematic review. JAMA; Nov 2; 294(17): 2203-9

Moore, GF. Audrey, S. Barker, M. Bond, L. Bonell, C. Hardeman, W. Moore, L. O'Cathain, A. Tinati, T. Wight, D. Baird, J. (2015) Process evaluation of complex interventions: Medical Research Council Guidance. BMJ. 350: 1258. Doi: 10.1136/bmj.h1258

Moore, GF. Audrey, S. Barker, M. Bond, L. Bonell, C. Cooper, C. Hardeman, W. Moore, L. O'Cathain, A. Tinati, T. Wight, D. Baird, J. (2014) Process evaluation in complex public health intervention studies: the need for guidance. Journal of Epidemiology Community Health; 68 (2): 101 – 102. Doi: 10.1136/jech-2013-202869

Moore, CG. Carter, RE. Nietert, PJ. Stewart, PW. (2011) Recommendations for planning pilot studies in clinical and translational research. Clinical and Translational Science; 4 (5): 332 – 337. Doi: 10.1111/j.1752-8062.2011.00347.x

Morris, T. Wan Hee, S. Stallard, N. Underwood, M, Patel, S. (2015) Can We Convert Between Outcome Measures of Disability for Chronic Low Back Pain? SPINE Volume 40, (10); 734-739

Mousavi, S. J. Parnianpour, M. Mehdian, H. Montazeri, A. Mobini, B. (2006) The Oswestry Disability Index, the Roland-Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: translation and validation studies of the Iranian versions. Spine (Phila Pa 1976); 31 (14): E454 – 459

Munnangi, S. Angus, LD. (2019) Placebo effects. StatPearls Publishing. https://www.ncbi.nim.nih.gov/books/NBK513296/

Musculoskeletal Association of Chartered Physiotherapists (MACP)

(2017) https://macpweb.org/home/index.php?p=366

Muller, T. (2009) Anti-inflammatory and analgesic electrotherapy. Evidence in rheumatology? Z Rheumatology; 68 (7): 532 -533. Doi: 10.1007/s00393-009-0444-4

Muller, R. Giles, LGF. (2005) Long-term follow-up of a randomised clinical trial assessing the efficacy of medication, acupuncture, and spinal manipulation for chronic mechanical spinal pain syndrome. Journal of Manipulative and Physiological Therapeutics; 28 (1): 3 – 11. Doi: 10.1016/j.jmpt.2004.12.004

Murnion, BP. (2018) Neuropathic pain: current definition and review of drug treatment. Australian Prescriber; 41 (3): 60 – 63. Doi: 10.18773/austprescr.2018.022

Mychaskiw, MA. Thomas, J. (2002) Direct costs of back pain in the United States: A national estimate. Value Health; 5 (6): 508

National Health Committee (NHC) (2004) National Advisory Committee on Health and Disability, Accident Rehabilitation and Compensation Insurance Corporation. New Zealand Acute Low back pain Guide. Wellington, New Zealand

NIHR – National Institute for Health Research (NIHR) (2017) NIHR research for patient benefit (RfPB) programme guidance on applying for feasibility studies. https://www.nihr.ac.uk/funding-and-support/documents/funding-for-research-studies/research-programmes/RfPB/Guidance%20Documents/Guidance_on_feasibility_studies.pdf

National Institute for Health Research (NIHR) (2017b) NIHR Clinical Trials Toolkit. www.ct-toolkit.ac.uk/routemap/safety-reporting/

National Institute for Health Research (NIHR) (2015) Evaluation, Trials and Studies; Glossary. http://www.nets.nihr.ac.uk/glossary/feasibility-studies.

Negrini, S. Giovabboni, S. Minozzi, S. Barneschi, G. Bonaiuti, D. Bussotti, A. D'Arienzo, M. DiLorenzo, N. Mannoni, A. Mattioli, S. Modena, V. Padua, L. Serafini, F. Violante, FS. (2006) Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines. Eura Medicophys; 42: (2) 151 – 170

NephJC. (2017) The method to the madness: Understanding factorial design trials. www.nephjc.com/news/2017/11/24/the-method-to-the-madness-understanding-factorial-design-trials

Neuhauser, H. Ellert, U. Ziese, T. (2005) Chronic back pain in the general population in Germany 2002/2003: prevalence and highly affected population groups. Gesundheitswesen 67: 685-693

New South Wales Therapeutic Assessment Group Inc (NSW TAG) (2002) Low back pain. Darlinghurst, NSW

NHS Digital (2017) Information and Technology for better health care. NHS digital; data and publications. https://digital.nhs.uk/article/191/Find-data-and-publications

NIAO - Northern Ireland Audit Office (2008) Management of sickness absence in the Northern Ireland civil service. Report by the Controller and Auditor General NIA 132/07-08, 22 May 2008. Norwich: TSO (The stationary office): www.niauditoffice.gov.uk/pubs/Absence/Absence-final.pdf

NICE - National Institute for Health and Care Excellence (2019) Consultation on improving how patients and the public can help develop NICE guidance and standards. www.nice.org.uk/about/nice-communities/nice-and-the-public/public-involvement/consultation

NICE (2018) 2018 Surveillance (exceptional review) of low back pain and sciatic in over 16s: assessment and management (NICE guideline NG59) www.nice.org.uk/guidance/ng59/resources/2018-surveillance-exceptional-review-of-low-back-pain-and-sciatic-in-over-16s-

assessment-and-management-NICE-guideline-NG59-6594317533

NICE (2017) Low back pain and Sciatica overview. London. National Institute for Health and Clinical Excellence (NICE) February. http://pathways.nice.org.uk/pathways/low-back-pain-and-sciatica

NICE (2016) Low back pain and sciatica in over 16s: assessment and management; Assessment and non-invasive treatments. NICE guideline NG59; Methods, evidence and recommendations; November. National Guideline Centre

NICE (2016b) Low back pain and sciatica in the overs 16s: assessment and management. NICE guideline. Published: 30 November 2016. Nice.org.uk/guidance/ng59

National Institute for Health and Care Excellence (NICE) (2009) Low Back Pain: Early Management of Persistent Non-Specific Low Back Pain. NICE Clinical Guideline 88. London; National Institute for Health and Clinical Excellence (NICE)

National Institute for Health and Care Excellence (NICE) (2009b) Low Back Pain: Costing Report: Implementing NICE Guidance. NICE Clinical Guideline 88. London; National Institute for Health and Clinical Excellence (NICE)

National Institute for Health and Care Excellence (NICE) (2009c) Quick Reference Guide: Low Back Pain; Early Management of Persistent Non-Specific Low Back Pain. NICE Clinical Guideline 88. London; National Institute for Health and Clinical Excellence (NICE)

National Institute for Health and Care Excellence (NICE) (2006) The Guidelines Manual. London: NICE, April. www.nice.org.uk/guidelinesmanual

Newman, T. (2017) Placebos: The power of the placebo effect. Medical News Today; https://www.medicalnewstoday.com/articles/306437.php

Nicolian, S. Butel, T. Gambotti, L. Durand, M. Filipovic-Pierucci, A. Mallet, A. Kone, M. Durand-Zaleski, I. Dommergues, M. (2019) Cost-effectiveness of acupuncture versus standard care for pelvic and low back pain in pregnancy: a randomized controlled trial. PLOS.ONE. https://doi.org/10.1371/journal.pone.0214195

Nielens H, Van Zundert J, Mairiaux P, Gailly J, Van Den Hecke N, Mazina D, et al. Chronic low back pain. Good Clinical practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE); (2006). KCE reports 48 C (D/2006/10.273/71)

Nijs, J. Apeldoorn, A. Hallegraeff, H. Clark, J. Smeets, R. Malfliet, A. Girbes, EL. De Kooning, M. Ickmans, K. (2015) Low back pain: Guidelines for the clinical classification of predominant neuropathic, nociceptive, or central sensitization pain. Pain Physician; 18: 333 – 346

Nishiwaki, M. Takayama, M. Yajima, H. Morihiro, N. Park. J. Kong, J. Takakura, N. (2018) A double-blind study on acupuncture sensations with Japanese style of acupuncture: Comparison between penetrating and placebo needles. Evidence-based Complementary and Alternative Medicine; 8128147: Doi: 10.1155/2018/8128147

Nordeman, L. Thorselius, L. Gunnarsson, R. Mannerkorpi, K. (2016) Predictors for future activity limitation in women with chronic low back pain consulting primary care: a 2-year prospective longitudinal cohort study. BMJ Open; 7 (6): Doi: 10.1136/bmjopen-2016-013974

Norvella, S. (2015) An industry of worthless acupuncture studies. Science-Based Medicine; https://sciencebasedmedicine.org/an-industry-of-worthless-acupuncture-studies/

Odole, AC. Akinpelu, AA. Adekanla, BA. Obisanya, OB. (2011) Economic burden of low back pain on patients seen at outpatient physiotherapy clinics of secondary and tertiary health institutions in Ibadan. Journal of the Nigeria Society of Physiotherapy; 18 (19): 43 – 48

Office of National Statistics (2017) People suffering with back pain 2015 to 2016: Freedom of information (FOI). Release date: 7th March 2017. https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/peoplesufferingwithbackpain2015to2016

O'Connell, NE. Cook, CE. Wand, BM. Ward, SP. (2017) Clinical guidelines for low back pain: A critical review of consensus and inconsistencies across three major guidelines. Best Practice & Research Clinical Rheumatology, http://dx.doi.org/10.1016/j.berh.2017.05.001

O'Sullivan, P. Lin, I. (2014) Acute low back pain. Beyond drug therapies. Pain Management Today; 1 (1) 8 – 14 http://www.pain-ed.com/wp-content/uploads/2014/02/Osullivan-and-Lin-Pain-management-today-2014.pdf

Oxford English dictionary (2015) Oxford: Oxford University Press, http://www.oxforddictionaries.com/definition/english/biased

Oxman, AD. Guyatt, GH. (1991) Validation of an index of the quality of review articles. Journal of Clinical Epidemiology; 44 (11): 1271 – 1278

Paavola, M. Malmivaara, A. Taimela, S. Kanto, K. Inkinen, J. Kalske, J.Sinisaari, I. Savolainen, V. Ranstam, J. Jarvinen, TLN. (2018) Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial. BMJ; 362: Doi: https://doi.org/10.1136/bmj.k2860

Pain Community Centre (PCC) (2010) Low back pain: the problem. http://www.paincommunitycentre.org/article/low-back-pain-problem

Pan, CX. Morrison, RS. Ness, J. Fugh-Berman, A. Leipzig, RM. (2000) Complementary and Alternative Medicine in the Management of Pain, Dyspnea, and Nausea and Vomiting Near the End of Life: A Systematic Review. Journal of Pain and Symptom Management; 20 (5): 374 – 387

Pannucci, CJ. Wilkins, EG. (2010) Identifying and avoiding bias in research. Plastic Reconstructive Surgery; 126 (2): 619 – 625

Papageorgiou, AR. (1991). Review of UK data on the rheumatic diseases - low back pain. British Journal of Rheumatology 30: 208-210

Parsons, S. Ingram, M. Clarke-Cornwell, AM. Symmons, DP. (2011) A Heavy burden. The occurrence and impact of musculoskeletal conditions in the United Kingdom today. The University of Manchester

Parsons, G. (2006) Acupuncture was better than no acupuncture but did not differ from minimal (sham) acupuncture for chronic low back pain at 8 weeks. Evidence Based Nursing; 9 (4): 111

Parthan, A. Evans, CJ. Le, K. (2006) Chronic Low back pain: epidemiology, economic burden and patient-reported outcomes in th USA. Expert Review of Pharmacoeconomics & Outcomes research; 6 (3): 359 – 369. Doi: org/10.1586/14737167.6.3.359

Pawar, SG. Ramani, PS. Prasad, A. Dhur, A. Babhulkar, SS. Bahurupi.YA. (2017) Software version of Roland Morris disability questionnaire for outcome assessment in low back pain. Neurological Research; 39 (4): 292 – 297. Doi: 10.1080/01616412.2017.1297555

Payakachatt, N. Ali, MM. Tilford, M. (2016) Can the EQ-5D detect a meaningful change? A systematic review. Parmacoeconomics; 33 (11): 1137 – 1154. Doi: 10.1007/s40273-015-0295-6

Petit, A. Mairiaux, P. Desarmenien, A. Meyer, JP. Roquelaure, Y. (2016) French good practice guidelines for management of the risk of low back pain among workers exposed to manual material handling: Hierarchical strategy of risk assessment of work situations. Work; 53 (4): 845 – 850. Doi: 10.3233/WOR-162258

Pezzilli, R. Morselli-Labate, AM. Frulloni, L. Cavestro, GM. Ferri, B. Comparato, G. Gullo, L. Corinaldesi, R. (2006) The quality of life in patients with chronic pancreatitis evaluated using the SF-12 questionnaire: A comparative study with the SF-36 questionnaire. Digestive and Liver Disease; 38 (2): 109 -115. https://doi.org/10.1016/j.dld.2005.09.015

Philippine Academy of Rehabilitation Medicine (PARM) (2011) Clinical practice guidelines on the diagnosis and management of low back pain. http://parm.com.ph/wp-content/uploads/2016/09/PARM-Low-Back-Pain-CPG-2011-1.pdf

Phillips, CJ. (2006) Economic burden of chronic low back pain. Expert Review of Pharmacoeconomics & outcome research; 6 (5): Doi: org/10.1586/14737167.6.5.591

Picavet, H. Hoeymans, N. (2004) Health related quality of life in multiple musculoskeletal diseases: SF-36 and EQ-5D in the DMC3 study. Annals of Rheumatic Diseases; 63 (6): 723 – 729. Doi: 10.1136/ard.2003.010769

Pillastrini, P. Gardenghi, I. Bonetti, F. Capra, F. Guccione, A. Mugnai, R. Violante, FS. (2012) An updated overview of clinical guidelines for chronic low back pain management in primary care. Joint Bone Spine; 79 (2): 176 – 185. Doi: 10.1016/j.jbspin.20112.03.019

Pillay, S. (2016) The psychology of low back pain. Harvard Health Blog, Harvard Health Publishing, Harvard Medical School; https://www.health.harvard.edu/blog/psychology-low-back-pain-201604259537

Pincus, T. McCraken, LM. (2013) Psychological factors and treatment opportunities in low back pain. Best Practice & Research Clinical Rheumatology; 27 (5): 625 – 635 https://doi.org/10.1016/j.berh.2013.09.010

Pocock, SJ. (1983) Clinical Trials: a Practical approach. Chichester: Wiley

PCRG - Preference Collaborative Review Group (PCRG) (2008). Patients' preferences within randomised trials: systematic review and patient level meta-analysis BMJ 337: a1864

Price, C. Hoggart, B. Olukoga, I. Williams, A. Bottle, A. (2012) National Pain Audit final report. The British Pain Society

Priedane, E. Phillips, G. Brazier, J. Johnson, CD. Gomes-Faria, R. Kumar, A. (2018) Limitations of EQ5D instrument in the assessment of quality of life in chronic rare diseases – a case study from FC. AKCEA Therapeutics; ISPOR, Europe, Barcelona, 10 – 14th November 2018 https://tools.ispor.org/research_pdfs/60/pdffiles/PSY197.pdf

PROSPERO (2019) International prospective register of systematic reviews. (July accessed online); http://www.crd.york.ac.uk/prospero

Qaseem, A. Wilt, TJ. McLean, RM. Forciea, MA. (2017) Noninvasive treatments for acute, sub-acute and chronic low back pain: A clinical practice guideline from the American College of Physicians. Annals of Internal Medicine; 166 (7): 514 – 530. Doi: 10.7326/M16-2367

Qaseem, A. Snow, V. Owens, DK. Shekelle, P. Clinical guidelines committee of the American Pain Society. (2010) The development of clinical practice guidelines and guidance statements of the American College of Physicians: Summary of methods. Annals of Internal Medicine; 153 (3): 194 – 199. Doi: 10.7326/0003-4819-153-3-201008030-00010

Qiu, K. Jing, M. Sun, R. Yang, J. Liu, X. He, Z. Yin, S. Lan, Y. Cheng, S. Gao, F. Liang, F. Zeng, F. (2016) The status of quality control in acupuncture-neuroimaging studies. Evidence-based Complementary and Alternative Medicine; 3685785: Doi: 10.1155/2016/3685785

Ratcliffe, J. Thomas, KJ. MacPherson, H. Brazier, J. (2006) A randomised controlled trial of acupuncture care for persistent low back pain: cost effectiveness analysis. BMJ; 333 (7569): 626. Doi: 10.1136/bmj.38932.806134.7C

Relton, C. et al. (2nd Twics symposium collaborators) (2017) The ethics of 'Trials within Cohorts' (TwiCs): 2nd international symposium. Meeting abstract. Trials; 18 (2): 244. Doi: 10.1186/s13063-017-1961-0

Relton, C. (2016) Trials within cohorts. Use of the TwiCs design.

https://www.twics.global/use-of-the-design

Relton, C. Thomas, K. Nicholl, J. Uher, R. (2015) Review of an innovative approach to practical trials: the 'cohort multiple RCT' design. Trials; 16 (2): 114. Doi: 10.1186/1745-6215-16-S2-P114

Relton, C. Torgerson, D. O'Cathain, A. Nicholl, J. (2010) Rethinking pragmatic randomized controlled trials: introducing the "cohort randomised controlled trial" design. British Medical Journal; 340: c1066

Relton, C. (2009) A new design for pragmatic randomised controlled trials: a 'Patient Cohort' RCT of treatment by a homeopath for menopausal hot flushes. URI: http://etheses.whiterose.ac.uk/id/eprint/6644

Reenen, MV. Janssen, B. (2015) EQ-5D-5L User guide. Basic information on how to use the EQ-5D-5L instrument. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf

Resnik, L. Dobrzykowsk, E. (2003) Guide to outcome measurement for patients with low back pain syndromes. Journal of Orthopaedic & Sports Physical Therapy; 33 (6): 307 – 318

Reynolds JA, Bland JM, MacPherson H. (2008) Acupuncture for irritable bowel syndrome: an exploratory randomised controlled trial. Acupuncture in Medicine. 26 (1): 8 – 16

Ricker, BA. Seymour, KM. Arslan, AR. Slee, A. McBride R. Probstfield, JL. (2015) Evaluating the effects of therapy masking. Applied Clinical Trials. March 13

Riksman, JS. Williamson, OD. Walker, BF. (2011) Delineating inflammatory and mechanical sub-types of low back pain: a pilot survey of fifty low back pain patients in a chiropractic setting. Chiropractic & Manual Therapies; 19: 5

Rizzo, JA. Berger, ML. Abbott, TA. (1998) The labor productivity effects of chronic backache in the United States. Medical Care; 36 (10): 1471 -

Rodgers, S. Sbizzera, I. Cockayne, S. Fairhurst, C. Lamb, SE. Vernon, W. Watson, J. Hewitt, C. Torgerson, D. (2018) A nested randomised controlled trial of a newsletter and Post-it note did not increase postal questionnaire response rates in a falls prevention trial. F1000Research; 7:1083. Doi: 10.12688/f1000research.14591.1

Roland, M; Fairbank, J (2000) The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine; 25 (24): 3115 – 3124

Roland, M. Morris, R. (1983) A Study of the Natural History of Back Pain: Part I: Development of a Reliable and Sensitive Measure of Disability in Low-Back Pain. Spine; 8 (2): 141 - 4

Ross, S. Grant, AM. Counsell, C. Gillespie, WJ. Russell, IT. Prescott, R. (1999) Barriers to participation in randomized controlled trials: A systematic review. Journal of Clinical Epidemiology; 52: 1143 – 1156

Rossignol, M. Rozenberg, S. Leclerc, A. (2009) Epidemiology of low back pain: what's new? Joint Bone Spine; Dec 76 (6): 608 – 613. Doi: 10.1016/j.jbspin.2009.07.003

Rossignol, M. Arsenault, B. Dionne, C. Tousignant, M. Truchon, M. Arsenault, B. Allard, P. Cote, M. Neveu, A. (2007) Clinic on low-back pain in interdisciplinary practice (CLIP) guidelines. Implementation Science; 2: 36. http://www.santpub-mtl.qc.ca/clip

Rothwell, JC. Julious, SA. Cooper, CL. (2018) A study of target effect sizes in randomized controlled trials published in the Health Technology Assessment journal. Trials; 19: 544 http://doi.org/10.1186/s13063-018-2886-y

Rubinstein, SM. Van Eekelen, R. Oosterhuis, T. de Boer, MR. Ostelo, RWJG. van Tulder, MW. (2014) The risk of bias and sample size of trial of spinal manipulative therapy for low back and neck pain: Analysis and

recommendations. Journal of Manipulative and Physiological Therapeutics; 37 (8): 523 – 541. Doi: 10.1016/j.jmpt.2014.07.007

Rubinstein, SM. Terwee, CB. Assendelft, WJ. de Boer, MR. van Tulder, MW. (2012) Spinal manipulative therapy for acute low-back pain. Cochrane Database Systematic Reviews; September 12; 9: CD008880. Doi: 10.1002/14651858. CD008880.pub2

Rubinstein, SM. van Middelkoop, M. Assendelft, WJ. de Boer, MR. van Tulder, MW. (2011) Spinal manipulative therapy for chronic low-back pain. Cochrane Database Systematic Reviews. Feb 16; (2): CD008112. Doi: 10.1002/14651858. CD008112.pub2

Ruddock, JK. Sallis, H. Ness, A. Perry, RE. (2016) Spinal manipulation Vs sham manipulation for nonspecific low back pain: a systematic review and meta-analysis. Journal of Chiropractic Medicine; 15 (3): 165 – 183. Doi: https://doi.org/10.1016/j.jcm.2016.04.014

Sadeghian, F. Hosseinzadeh, S. Aliyari, R. (2014) Do psychological factors increase the risk for low back pain among nurse? A comparing according to cross-sectional and prospective analysis. Safety and Health at Work; 5 (1): 13 -16 https://doi.org/10.1016/j.shaw.2013.11.004Get rights and content

Sahin, N. Albayrak, I. Durmus, B. Ugurlu, H. (2011) Effectiveness of back school for treatment of pain and functional disability in patients with chronic low back pain: a randomized controlled trial. Journal of Rehabilitation Medicine; 43: 224 – 229

Salter, GC. Roman, M. Bland, JM. Macpherson, H. (2006) Acupuncture for chronic neck pain: a pilot for a randomized controlled trial. BMC Musculoskeletal Disorders. 7: 99

Salerno, SM. Browning, R. Jackson, JL. (2002) The effect of antidepressant treatment on chronic back pain. Archives of Internal Medicine; 162: 19 – 24

Santilli, V. Beghi, E. Finucci, S. Chiropractic manipulation in the treatment of acute low back pain and sciatica with disc protrusion: a randomized double-blind clinical trial of active and simulated spinal manipulations. Spine; 6 (2): 131 – 137. Doi: 10.1016/j.spinee.2005.08.001

SAR – The Society for Acupuncture Research (2018) How to identify and assess well-designed trial research. https://www.acupunctureresearch.org/identifying-quality-research-studies

Savigny, P. Kuntze, S. Watson, P. Underwood, M. Guideline Development Group. (2009) Low back pain: early management of persistent non-specific low back pain. National Collaborating Centre for Primary Care (NCCPC) and Royal College of General Practitioners (RCGP), London, UK. BMJ; 4: 388: b1805. Doi: 10.1136/bmj.b1805

Savulescu, J. Wartolowska, K. Carr, A. (2016) Randomised placebo-controlled trials of surgery: ethical analysis and guidelines. Journal of Medical Ethics; 42 (12): 776 – 783. Doi: 10.1136/medethics-2015-103333

Schober, P. Bossers, S. Schwarte, Lothar A. (2018) Statistical Significance Versus Clinical Importance of Observed Effect Sizes: What Do *P* Values and Confidence Intervals Really Represent? Anaesthesia & Analgesia: March 2018 - Volume 126 - Issue 3 - p 1068-1072. Doi: 10.1213/ANE.0000000000002798

Schümenann, H. Brożek, J. Guyatt, G. Oxman, A. (2013) GRADE Handbook. Introduction to GRADE. Updated October 2013. www.gdt.gradepro.org

Scottish Intercollegiate Guidelines Network (SIGN) (2013) Management of Chronic Pain. A National Clinical Guideline. (Rockville, MD: Agency for healthcare research and quality) Edinburgh: SIGN publication no. 136. ISBN 978 1 909103 17 7 http://www.sign.ac.uk

Sedgwick, P. (2015) Understanding the Hawthorne effect. BMJ; 351. Doi:

10.1136/bmj.h4672

Sedgwick, P. (2014a) Explanatory trials versus pragmatic trials. BMJ; 349. Doi: 2014;349:g6694

Sedgwick, P. (2014) What is a factorial study design? BMJ; 349 (1): g5455. Doi: 10.1136/bmj.g5455

Sedgwick, P. (2012) Randomised controlled trials with full factorial designs. BMJ; 345: Doi: https://doi.org/10.1136/bmj.e5114

Sedgwick, P. (2012b) The Hawthorne Effect. BMJ; 344: Doi: 10.1136/bmj.d8262

Senn, S. (1993) Cross-over. Trials in clinical research. Wiley, Chichester, UK. ISBN 0-471-93493-3

Senna, MK. Machaly, SA. (2011) Does maintained spinal manipulation therapy for chronic nonspecific low back pain result in better long-term outcome? Spine; 36 (18): 1427 – 1437. Doi: 10.1097/BRS.0b013e3181f5dfe0

Sherman, KJ. Cherkin, DC. Ichikwa, L. Avins, AL. Barlow, WE. Khalsa, PS. Deyo, RA (2009) Characteristics of patients with chronic low back pain who benefit from acupuncture. BMC Musculoskeletal Disorders; 10:114. Doi: 10.1186/1471-2474-10-114

Sharim, M. Cifuentes, M. Willetts, JL. Marucci-Wellman, HR. Pransky, G. (2017) Regional socioeconomic disparities in outcomes for workers with low back pain in the United States. American Journal of Industrial Medicine; 60 (5): 472 – 483. Doi: 10.1002/ajim.22712

Shiel, WC. (2018) Medical definition of cytokine. MedicineNET. https://www.medicinenet.com/script/main/art.asp?articlekey=11937

Shih, ML. Costi, JM. Teixeira, JEM. (2008) Sham acupuncture is not a placebo. Archives of Internal Medicine; 168 (9): 1011. Doi: 10.1001/archinte.168.9.1011-b

Sibbald, B. Roberts, C. (1998) Understanding controlled trials. Crossover trials. British Medical Journal (BMJ), 316: 1719

Silverman, WA. (1998) Where's the evidence: Debates in Modern Medicine. Oxford: Oxford University Press

Singh, SP. Tuomainen, H. Girolamo, G. Maras, A. Santosh, P. McNicholas, F. Schulze, U. Purper-Ouakil, D. Tremmery, S. Franić, T. Madan, J. Paul, M. Verhulst, FC. Dieleman, GC. Warwick, J. Wolke, D. Street, C. Daffren, C. Tah, P. Griffin, J. Canaway, A. Signorini, G. Gerritsen, S. Adams, L. O'Hara, L. Aslan, S. Russet, F. Daidović, N. Tuffrey, A. Wilson, A. Gatherer, C. Walker, L. (MILESTONE Consortium) (2017) Protocol for a cohort study of adolescent mental health service users with a nested cluster randomised controlled trial to assess the clinical and cost-effectiveness of managed transition in improving transitions from child to adult mental health services (the MILESTONE study) BMJ Open; 7: e016055. Doi: 10.1136/bmjopen-2017-016055

Singh, S. (2017) How to conduct and interpret systematic reviews and meta-analysis. Clinical and Translational Gastroenterology; 8 (e93): 1 – 5. Doi: 10.1038/ctg.2017.20

Singh, A. Gnanalingham, K. Casey, A. Crockard, A. (2006) Quality of life assessment using the short form-12 (SF-12) Questionnaire in patients with cervical spondylotic myelopathy. Spine; 31 (6): 639 - 643

Sivasubramaniam, V. Patel, HC. Ozdem, BA. Papadopolous, MC. (2015) Trends in hospital admissions and surgical procedures for degenerative lumbar spine disease in England: a 15-year time-series study. BMJ Open; 5: e009011. Doi: 10.1136/bmjopen-2015-009011 http://dx.doi.org/10.1136/bmjopen-2015-009011

Skivington, K. Matthews, L. Craig, P. Simpson, S. Moore, L. (2018) Developing and evaluating complex interventions: updating Medical Research Council guidance to take account of new methodological and

theoretical approaches. The Lancet; 392 (S2). Doi: https://doi.org/10.1016/S0140-6736(18)32865-4

Skrabanek, P. (2003) Nonsensus consensus. The Lancet; 335 (8703): 1446 – 1447. Doi: 10.1016/0140-6736(90)9146-R

Slater, SL. Ford, JJ. Richards, MC. Taylor, NF. Surkitt, LD. Hahne, AJ. (2012) The effectiveness of sub-group specific manual therapy for low back pain: a systematic review. Manual Therapy; 17 (3): 201 – 12. Doi: 10.1016/j.math.2012.01.006

Small, DS. Volpp, KG. Rosenbaum, PR. (2011) Structured testing of 2 x 2 factorial effects: an analytic plan requiring observations. The American Statistician; 65 (1): 11 – 15.

Smart, KM. Blake C, Staines A, Doody C. (2010) Clinical indicators of "nociceptive", 'peripheral neuropathic' and 'central' mechanisms of musculoskeletal pain. A Delphi survey of expert clinicians. Manual Therapy; 15: 80 – 87

Smeets, R. Köke, A. Chung-Wei, L. Manuela, F. Demoulin, C. (2011) Measures of function in low back pain/disorders: Low Back Pain Rating Scale (LBPRS), Oswestry Disability Index (ODI), Progressive isoinertial Lifting Evaluation (PILE), Quebec Back Disability Scale (QBPDS) and Roland-Morris Disability Questionnaire (RDQ). Arthritis and Care Research; 63 (S11): S158 –S173. https://doi.org/10.1002/arc.20542

Snider, KT. Johnson, JC. Snider, EJ. Degenhardt, BF. (2008) Increased incidence and severity of somatic dysfunction in subjects with chronic low back pain. The Journal of the American Osteopathic Association; 108 (8): 372 – 378

Spanish Back Pain Research Network, Spain (2005) Guia de practica clinica. Lumbalgia Inespecifica. Version espnola de la Guia de Practica Clinica del Programa Europeo COST B13

Spanjer, J. Groothoff, JW. Brouwer, S. (2011) Instruments used to assess

functional limitations in workers applying for dis- ability benefit: a systematic review. Disability Rehabilitation; 33 (23-24): 2143 - 2150

Spieth, PM. Kubasch, AS. Penzlin, AI. Illigens, BM. Barlinn, K. Siepmann, T. (2016) Randomized controlled trials – a matter of design. Neuropsychiatric Disease and Treatment; 12: 1341 – 1349. Doi: 10.2147/NDT.S101938

SPRINTER (2005) Randomised Trial of physiotherapy for neck pain British Medical Journal (BMJ); 8: 330 (7482): 75. Doi: 10.1136/bmj.38286.493206.82

Standaert, CJ. Friedly, J. Erwin, MW. Lee, MJ. Rechtine, G. Henrikson, NB. Norvell, DC. (2011) Comparative effectiveness of exercise, acupuncture, and spinal manipulation for low back pain. Spine; 36 (21 Supplement): S120-S130

Stein, C. Machelska, H. Binder, W. Schafer, M. (2001) Peripheral opioid analgesia. Current Opinion in Pharmacology; 1 (1): 62 - 65

Stevens, ML. Lin, CCW. Maher, CG. (2016) The Roland Morris Disability Questionnaire. Journal of Physiotherapy; 62 (2): 116. https://doi.org/10.1016/j.phys.2015.10.003

Stevens, A. Abrams, KR. Brazier, J. Fitzpatrick, R. Lilford, RJ. (2001) The Advanced handbook of methods in evidence-based healthcare. Sage Publications, London ISBN 0 7619 6144 5

Stewart, L. Moher, D. Shekelle, P. (2012) Why prospective registration of systematic reviews makes sense. Systematic Reviews; 1 (7): Doi: 10.1186/2046-4053-1-7

Stone, PW. (2002) Popping the (PICO) question in research and evidence-based practice. Applied Nursing; 15 (3): 197 – 8

Stratford, PW. Riddle, DL. (2016) A Roland Morris disability questionnaire target value to distinguish between functional and

dysfunctional states in people with low back pain. Physiotherapy Canada; 68 (1): 29 – 35. Doi: 10.3138/ptc.2014-85

Stratford, PW. Binkley, J. Solomon, P. Finch, E. Gill, C. Moreland, J. (1996) Defining the minimum level of detectable change for the Roland-Morris questionnaire. Physical Therapy; 76 (4): 359 - 365

Stratford, PW. Binkley, J. Soloman, P. Gill, C. Finch, E. (1994) Assessing change over time in patients with low back pain. Physical Therapy; 74 (6): 528 – 533. Doi: 10.1093/ptj/74.6.528

Streiner, DL. Norman, GR. (2008) Health measurement scales. A practical guide to their development and use. Fourth Edition. Oxford University Press Inc: New York. ISBN-13: 978-0199231881

Streiner, DL. Norman, GR. (1989) Health measurement scales. A practical guide to their development and use. Second Edition. Oxford University Press Inc: New York.

Strite, SA. (2010) Sources of bias in randomised controlled trials II. BMJ; 341: Doi: 10.1136/bmj.c7053

Strietberger, K. Kleinhenz, J. (1998) Introducing a placebo needle into acupuncture research. Lancet; 352 (9125): 364 – 365. Doi: 10.1016/S0140-6736(97)10471-8

Sydes, M, R. Parmar, M. James, N, D. Clarke, N, W. Dearnaley, D, P. Mason, M, D. Morgan, RC. Sanders, K. Royston, P. (2009) Issues in applying multi-arm multi-stage methodology to a clinical trial in prostate cancer. The MRC STAMPEDE trial. Trials, 10:39 http://www.trialsjournal.com/content/10/1/39

Tatta, J. (2017) The role of central sensitization in chronic back pain. 31 January. http://www.drjoetatta.com/central-sensitization-chronic-back-pain/

The Association of Chartered Physiotherapists in Sport and Exercise Medicine (ACPSEM) (2017) https://www.physiosinsport.org

Thelin, A. Holmberg, S. Thelin, N. (2008) Functioning in neck and low back pain from a 12-year perspective: a prospective population-based study. Journal of Rehabilitation Medicine; 40: 555 – 61

Thomas KJ, MacPherson H, Thorpe L, Brazier J, Fitter M, Campbell MJ, Roman M, Walters SJ, Nicholl J. (2006) Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. BMJ; Sep 23; 333 (7569): 623

Thorpe, KE. Zwarenstein, M. Oxman, AD. Trewwek, S. Furberg, CD. Altman, DG. Tunis, S. Bergel, E. Harvey, I. Magid, DJ. Chalkidou, K. (2009) A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. Canadian Medical Association Journal (CMAJ) 180 (10). Doi: 10.1503/cmaj.090523

Tilbrook, HE. Cox, H. Hewitt, CE. Kang'ombe, AR. Chuang, LH. Jayakody, S. et al. (2011) Yoga for chronic low back pain: a randomized trial. Annuals Internal Medicine; 155: 569e78

Tonosu, J. Oka, H. Higashikawa, A. Okazaki, H. Tanaka, S. Matsudaira, K. (2017) The associations between magnetic resonance imaging findings and low back pain: a 10-year longitudinal analysis. PLOS one; 12 (11): e0188056. Doi: 10.1371/journal.pone.0188057

TOP (Towards Optimized Practice) Program. (2017) Guidelines for the evidence-informed primary care management of low back pain; 3rd Edition. Institute of Health Economics. Alberta: Canada

TOP (Towards Optimized Practice) Program. (2015) Guidelines for the evidence-informed primary care management of low back pain: 2nd Edition, December. Institute of Health Economics. Alberta: Canada

TOP (Towards Optimized Practice) Program. (2009) Guidelines for the evidence-informed primary care management of low back pain. Institute

of Health Economics. Alberta: Canada

Torgerson, D J. Torgerson, C J (2008) Designing randomised controlled trials in Health, Education and Social Sciences, An introduction. Palgrave Macmillan. 117-118

Torgerson, DJ. (2001) Commentary: problems with randomised consent. British Medical Journal (BMJ) 322: 495

Torgerson, DJ. Kaber-Moffett, JA. Russell, IT. (1996) Patient preferences in randomised trials: threat or opportunity. Journal of Health Services Research and Policy; 1: 194 – 7

Torres-Rosas, R. Yehia, G. Pena, G. Mishra, P. Thompson-Bonilla, MR. Moreno-Eutimio, MA. (2014) Dopamine mediates vagal modulation of the immune system by electro-acupuncture. Nature Medicine; 20 (3): 291 -295

Toward Optimized Practice (TOP) (2015) Low back pain working group. Evidence-informed primary care management of low back pain: clinical practice guideline. Edmonton, AB Toward Optimized Practice; www.topalbertadoctors.org/cpgs/885801

Triano, JJ. McGregor, M. Hondras, MA. Brennan, PC. (1995) Manipulative therapy versus education programs in chronic low back pain. Spine; 20 (8): 948 - 955

van Tulder, MW. Becker, A. Bekkering, T. Breen, A. Teresa Gil del Real, M. Hutchinson, A. Koes, B Laerum, E. Malmivaara, A. and on behalf of the COST B13 Working Group on Guidelines for the Management of Acute Low Back Pain in Primary Care (2006) European guidelines for the management of acute low back pain in primary care. European Spine Journal; 15 (Suppl 2): S169 – S191. Doi: 10.1007/s00586-006-1071-2

van Tulder, MW. Koes, BW. Bouter, LM. (1997) Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. Spine; 22

(18): 2128 – 2156

Varatharajan, S. Chen, WS. (2012) Reliability and validity of EQ-5D in Malaysian population. Applied Research in Quality of Life; 7 (2): 209 – 221

Turk, SC. Melzack, R. (2011) *Handbook of pain assessment*, 3rd edition. New York; Guilford Press

Turk, DC. Swanson, KS. Tunks, ER. (2008) Psychological approaches in the treatment of chronic pain patients – when pills, scaples and needles are not enough. The Canadian Journal of Psychiatry; 53 (4): 213 – 223. Doi: 10.1177/070674370805300402

Uchida, S. Hotta, H. (2008) Acupuncture affects regional blood flow in various organs. Evidence-Based Complementary and Alternative Medicine; 5 (2): 145 -151. Doi: 10.1093/ecam/nem051

UKBEAM. United Kingdom back pain exercise and manipulation (UK BEAM) (2004) Randomised trial: effectiveness of physical treatments for back pain in primary care. British Medical Journal (BMJ); 329: 1377

Utah (2001) Interpreting the SF-12, Utah Health Status Survey, Utah Department of Health

Vale, CL. Rydzewska, LH. Rovers, MM. Emberson, JR. Gueyffier, F. Stewart, LA. (2015) Uptake of systematic reviews and meta-analysis based on individual participant data in clinical practice guidelines: descriptive study. BMJ; 350. Doi: https://doi.org/10.1136/bmj.h1088

Van der Roer, N Goossens, ME. Evers, SM. Van Tulder, MW. (2005) What is the most cost effective treatment for patients with low back pain? A systematic review. Best Practice Research in Clinical Rheumatology; 19 (4): 671 – 684

Van Tulder, M. Becker, A. Bekkering, T. Breen, A. Gil del Real, MT. Hutchinson, A. Koes, B. Laerum, E. Malmivaara, A. (2006) European

guidelines for the management of acute nonspecific low back pain in primary care. European Spine Journal; 15 (2): S169 – S191. Doi: 10.1007/s00586-006-1071-2

Van Tulder, MW. Ostelo, R. Vlaeyen, JWS. Linton, SJ. Morley, SJ. Assendelft, WJJ. (2000) Behavioral Treatment for Chronic Low Back Pain. Spine; 26 (3): 270 – 281

Van Wambeke, P. Desomer, A. Ailliet, L. Berquin, A. Demoulin, C. Depreitere, B. Dewachter, J. Dolphens, M. Forget, P. Fraselle, V. Hans, G. Hoste, Mahieu, G. Michelsen, J. Nielems, H. Orban, T. Parlevliet, T. Simons, E. Tobbackx, Y. Van Schaeybroeck, P. Van Zundert, J. Vandersyraten, J. Vlaeyen, J. Jonckheer, P. (2017) Low back pain and radicular pain: assessment and management. KCE Report 287: Good Clinical Practice. www.kce.fgov.be

Vas, J. Aranda, JM. Modesta, M. Benitez-Parejo, N. Herrera, A. Martinez-Barquin, DM. Aguilar, I. Sánchez-Araujo, M. Rivas-Ruiz, F. (2012) Acupuncture in patients with acute low back pain: a multicenter randomised controlled clinical trial. Pain: 153 (9): 1883 - 1889

Velengtas, P. Mohr, P. Messner, D. (2012) Making informed decisions: assessing the strengths and weaknesses of study designs and analytic methods for comparative effectiveness research. A briefing document for stakeholders. National Pharmaceutical Council

Vera, L. Zotelli, R. Grillo, CM. Bressiani, ML. Gil, B. Wada, RS. Sato, JE. De Sousa, MLR. (2018) Patterns of energy imbalance of the meridians in patients with temporo-mandibular dysfunction. Journal of Acupuncture and Meridian Studies; 11 (1): 1 – 6. Doi: 10.1016/j.jams.2017.11.002

Vianin, M. (2008) Psychometric and clinical usefulness of the Oswestry Disability Index. Journal of Chiropractic Medicine; 7 (4): 161 – 163. Doi: 10.1016/j.jcm.2008.07.001

Vickers, AJ. Young-Afat, DA. Ehdaie, B. Kim, SYH. (2018) Just-in-time

consent: The ethical case for an alternative to traditional informed consent in randomized trials comparing an experimental intervention with usual care. Clinical Trials; 15 (1) 3 – 8. Doi: 10.1177/1740774517746610

Vickers, AJ. Vertosick, EA. Lewith, G. MacPherson, H. Foster, NE. Sherman, KJ. Irnich, D. Witt, CM. Linde, K. on behalf of the Acupuncture Trialists Collaboration. (2018b) Acupuncture for chronic pain: Update of an individual patient data meta-analysis. The Journal of Pain; 19 (5): 455 – 474. Doi: 10.1016/j.jpain.2017.11.005

Vickers, A. Wilson, P. Kleijen, J. (2002) Acupuncture – Effectiveness Bulletin. BMJ Quality & Quality; 11 (1): 92 - 97

Vickers, A. Goyal, N. (1998) Do certain countries produce only positive results? A systematic review of controlled trials. Controlled Clinical Trials 19: 159 – 166

Viksveen, P. Relton, C. Nicholl, J. (2017) Depressed patients treated by homeopaths: a randomised controlled trial using the "cohort multiple randomised controlled trial" (cmRCT) design. Trials; 18 (1): 299. Doi: 10.1186/s13063-017-2040-2

Viswanathan, M. Berkman, ND. Dryden, DM. Hartling, L. (2013) Assessing risk of bias and confounding in observational studies of interventions or exposures. Methods Research Reports: Agency for Healthcare Research Quality (US): 13-EHC106-EF

Vos, T. Allen, C. Arora, M. Murray, CJL. et al. (637 authors) (2016) Global, regional, and national incidence, prevalence and years lived with disability for 310 diseases and injuries, 1990 – 2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet; 388 (10053): 1545 – 1602 Doi: 10.1016/S0140-6736(16)31678-6

Vrontou, S. Wong, AM. Rau, KK. Koerber, R. Anderson, D. (2013) Genetic identification of C fibres that direct massage-like stroking of hairy

skin in vivo. Nature; 493: 660 – 673 www.nature.com/atricles/nature11810

Waagen, G. Halderman, S. Cook, G. Lopez, D. Deboer, K. (1986) Short term trial of chiropractic adjustments for relief of chronic low back pain. Manual Medicine; 2: 63 - 67

Wager, TD. Atlas, LY. (2015) The neuroscience of placebo effects: connecting context, learning and health. Nature Reviews Neuroscience; 16: 403 – 418

Walker, BF. French, SD. Grant, W. Green, S. (2010) Combined chiropractic interventions for lo-back pain. Cochrane Database of Systematic Reviews; (4): CD005427. Doi: 10.1002/14651858.CD005427.pub2

Walker, B. (2000) The Prevalence of low back pain: a systematic review of the literature from 1966 to 1998. Journal of Spinal Disorders; 13: 205 – 217

Wand, BM. O'Connell, N. (2008) German acupuncture trials for chronic low back pain. Archives of Internal Medicine; 168 (9): 1011. Doi: 10.1001/archinte168.9.1011-a

Wang, Y. Zhao, J. Hao, D. (2017) Is sham acupuncture a real placebo: Sceptical for sham acupuncture. World Journal of Acupuncture – Moxibustion; 27 (2): 1 -5. Doi: 10.1016/S1003-5257(17)30110-1

Wang, Y. Gehringer, R. Mousa, SA. Hackel, D. Brack, A. Rittner, HL. (2014) CXCL10 controls inflammatory pain via opioid peptide-containing macrophages in electro-acupuncture. PLUS One; 9 (4): e94696. Doi: 10.1371/journal.pone.00094696

Wang, SM. Kain, ZNK. White, P. (2008) Acupuncture analgesia: 1. The scientific basis. Pain Medicine; 106 (2): 602 - 610

Ware, JE. Kosinski, MA. Keller, SD. (1995) SF-12: How to score the SF-12 physical and mental health summary scales. The Health Institute, New England Medical Centre, Boston, Massachusetts

Waterman, BR. Belmont, PJ. Schoenfeld, AJ. (2010) Low back pain in the United States: incidence and risk factors for presentation in the emergency setting. The Spinal Journal; 12: 63 – 70

Waterman, BR. Belmont, PJ. Schoenfeld, AJ. (2012) Low back pain in the United States and risk factors for presentation in the emergency setting. Spine; 12 (1): 63 – 70. Doi: 10.1016/j.spinee.2011.09.002

Watson, T. (2008) Electrotherapy: Evidence based practice. 12th Edition. Elsevier / Churchill Livingston. ISBN: 978-0-443-10179-3

Watson, T. (2000) The role of electrotherapy in contemporary physiotherapy practice. Manual Therapy; 5(3): 132 – 141

Weeks, J. (2016) Acupuncture group challenges back pain guidelines as "evidence-biased medicine". Acupuncture Now Foundation; www.theintegrativepractitioner.com/topics/news/acupuncture-group-challenges-back-pain-guideline-evidence-biased-medicine

Weijer, C. Goldstein, CE. Taljaard, M. (2017) TwiC or treat? Are trials within cohorts ethically defensible? Clinical Trials; 15 (1): 21 – 24. Doi: 10.1177/1740774517746622

White, A. Cummings, M. (2012) Inconsistent placebo effects in NICE's network analysis. Acupuncture in Medicine; 30: 364 – 365. Doi: 10.1136/acupmed-2012-010262

White, A. Editorial Board of Acupuncture in Medicine (2009) Western medical acupuncture: a definition. Acupuncture in Medicine; 27 (1): 33 – 35. Doi: 10.1136/aim.2008.000372

White, A (2006) The safety of acupuncture - evidence from the UK. Acupuncture in Medicine; 24: (suppl.) 53 – 57

White, A.R. (2002a). Acupuncture research methodology. In Lewith, GT. Jonas, WB. Walach, H. (eds.) Clinical Research in Complementary Therapies: Principles, Problems and Solutions. Churchill Livingstone: Edinburgh. 307 – 323

White, P. Lewith, G. Berman, B. Birch, S. (2002) Reviews of acupuncture for chronic neck pain: pitfalls in conducting systematic reviews. *Rheumatology*, 41, 1224 – 1231

White, A. Hayhoe, S. Hart, A. Ernst, E. (2001) Survey of adverse events following acupuncture (SAFA): a prospective study of 32,000 consultations. Acupuncture in Medicine; 19 (2), 84 – 92

Whitehead, AL. Julious, SA. Cooper, CL. Campbell, MJ. (2016) Estimating the sample size for a pilot randomised trial to minimize the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Statistical Methods in Medical Research; 25 (3): 1057 – 1073. Doi: 10.1177/0962280215588241

Wilson, ECF. (2015) A practical guide to value of information analysis. PharmacoEconomics; 33 (2): 105 – 121. Doi: https://doi.org/10.1007/s40273-014-0219-x

Wittes, J. (2002) Sample size calculations for randomized controlled trials. Epidemiologic Reviews; 24 (1)

Wong, JJ. Côté, P. Sutton DA. Randhawa, K. Yu, H. Varatharajan, S. Goldgrub, R. Nordin, M. Gross, DP. Shearer, HM. Carroll, LJ. Stern, PJ. AMeis, A. Southrest, D. Mior, S. Stupar, M. Varatharajan, T. Taylor-Vaisey, A. (2017) Clinical practice guidelines for the noninvasive management of low back pain: A systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. European Journal of Pain; 21 (2): 201 – 216. Doi: 10.1002/ejp.931

World Health Organization (WHO) (2018) Global, regional, and national incidence, prevalence, and years lived with disability for 345 diseases and

injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. GBD 2017 Disease and Injury incidence and prevalence collaborators. The Lancet; 392 (10159): 1789 – 1858. Doi: https://doi.org/10.1016/S0140-6736(18)32279-7

World Health Organization (WHO) (2014) WHO Handbook for guideline development. 2nd Edition. WHO Press ISBN: 97892 4 1548960

World Health Organisation (WHO) 2013. Priority diseases and reasons for inclusion: Low back pain. 6: 24. http://www.who.int/medicines/areas/priority_medicines/Ch6_24LBP.pdf

World Health Organisation (WHO) (2003). Low back pain. Bulletin of the World Health Organisation; 81: 671 – 676

World Health Organisation (WHO) (2002). Collection of WHO's documents for acupuncture. The World Federal of Acupuncture-Moxibustion Societies (WFAS). Beijing, China

World Health Organization (WHO). (1991) Traditional medicine and modern health care. Progress report by the Director-General. Geneva, chapter five (unpublished document A44/10)

Xiang, Y. He, J. Li, R. (2018) Appropriateness of sham or placebo acupuncture for randomized controlled trials of acupuncture for nonspecific low back pain: a systematic review and meta-analysis. Journal of Pain Research; 11: 83 – 94. Doi: 10.2147/JPR.S152743

Xu, M. Yan, S. Yin, X. Li, X. Gao, S. Han, R. Wei, L. Luo, W. Lei, G (2013) Acupuncture for chronic low back pain in long-term follow-up: a meta-analysis of 13 randomized controlled trials. The American Journal of Chinese Medicine; 41 (1): 1 – 9

Yang, M. Stufken, J. (2008) Optimal and efficient crossover designs for comparing test treatments to a control treatment under various models. Journal of Statistical Planning and Inference; 138 (1): 278 – 285. Doi: 10.1016/j.jspi.2007.05.005

Yelland, LN. Sullivan, TR. Voysey, M. Lee, KJ. Cook, JA. Forbes, AB (2015) Applying the intention-to-treat principle in practice: Guidance on handling randomisation errors. Clinical Trials; 12 (4) 418 – 423. Doi: 10.1177/1740774515588097

Yun, M. Shao, Y. Zhang, Y. He, S. Xiong, N. Zhang, J. Guo, M. Liu, D. Luo, Y. Guo, L. Yan, J. (2012) Hegu acupuncture for chronic low-back-pain: a randomized controlled trial. Journal of Alternative and complementary Medicine; 18: (20) 130 – 135

Zeng, Y. Chung, JW. (2015) Acupuncture for chronic nonspecific low back pain: An overview of systematic reviews. European Journal of Integrative Medicine; 7 (2) 94 – 107. Doi: 10.1016/j.eujim.2014.11.001

Zhang, LL. Alonso-Coello, P. Guyatt, GH. Yepes-Nuñez, JJ. Akl, EA. Hazlewood, G. Pardo-Hernandez, H. Etxeandia-Ikobaltzeta, I. Qaseem, A. Williams, JW. Tugwell, P.Flottro, S. Chang, Y. Zhang, Y.Mustafa, RA. Rojas, X. Schünermann, HJ. (2018) GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences – inconsistency, imprecision, and other domains. Journal of Clinical Epidemiology; 111: 83 – 93. Doi: https://doi.org/10.1016/j.jclinepi.2018.05.011

Zhang, LL. Alonso-Coello, P. Guyatt, GH. Yepes-Nuñez, JJ. Akl, EA. Hazlewood, G. Pardo-Hernandez, H. Etxeandia-Ikobaltzeta, I. Qaseem, A. Williams, JW. Tugwell, P.Flottro, S. Chang, Y. Zhang, Y.Mustafa, RA. Rojas, X. Schünermann, HJ. (2018) GRADE guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences-Risk of bias and indirectness. Journal of Clinical Epidemiology; 111: 94 – 104. Doi: https://doi.org/10.1016/j.jclinepi.2018.01.013

Zhang, LL. Chu, Q. Wang, S. Lai, H. Xie, BB. (2016) Is sham acupuncture as effective as traditional Chinese acupuncture? It's too early to say. Chinese Journal of Integrative Medicine; 22 (7): 483 – 489. https://doi.org/10.1007/s11655-016-2458-5

Zhang, CS. Tan, HY. Zhang, GS. Zhang, AL. Xue, CC. Xie, YM. (2015) Placebo devices as effective control methods in acupuncture clinical trials: A systematic review. PLoS ONE 10 (11): e0140825 https://doi.org/10.1371/journal.pone.0140825

Zhang, WT. Jin, Z. Luo, F. Zhang, L. Zeng, YW. Han, JS. (2004) Evidence from brain imaging with fMRI supporting functional specificity of acupoints in humans. Neuroscience; 354: 50 – 53

Zhu, D. Gao, Y. Chang, J. Kong, J. (2013) Placebo acupuncture devices: considerations for acupuncture research. Evidence based Complementary and Alternative medicine; 628907: Doi: 10.1155/2013/628907