The association between non-specific factors and response to primary care treatments for low back pain: a synthesis of evidence

Majid Artus

Doctor of Philosophy

2011

Keele University

Keele Graduate School

SUBMISSION OF THESIS FOR A RESEARCH DEGREE

Degree for which thesis being submitted

Title of thesis The association between non-specific factors and response to primary care treatments for low back pain: a synthesis of evidence

Date of submission 27th October 2011 Original registration date 1st June 2007

Name of candidate Majid Artus

Research Institute Primary Care and Health Sciences

Name of Lead Supervisor Professor Danielle van der Windt

DECLARATION by the candidate for a research degree

I certify that:

(a) the decision to submit this thesis has been taken following consultation with, and having received advice from, my academic and (where appropriate) non-academic (e.g. industrial) supervisor(s)

(b) The thesis being submitted for examination is my own account of my own research

(c) My research has been conducted ethically. Where relevant a letter from the approving body confirming that ethical approval has been given has been bound in the thesis as an Annex

(d) The data and results presented are the genuine data and results actually obtained by me during the conduct of the research

(e) Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis

(f) Where any collaboration has taken place with one or more other researchers, I have included within an 'Acknowledgments' section in the thesis a clear statement of their contributions, in line with the relevant statement in the Code of Practice (see Note overleaf).

(g) The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination

(h) Where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (if any), this has been identified and acknowledged in the thesis

(i) The thesis submitted is within the required word limit as specified in the Regulations

Total words in submitted thesis (including the text and footnotes, but excluding references and appendices) 74,817

Signature of candidate ……Date ……

Declaration

The literature searches and reviews were planned and conducted by myself with advice from a research information manager, Joanne Jordan. Inclusion of studies in the review and extraction of data from studies were conducted by myself and checked by my supervisors, Professor Danielle van der Windt and Dr Kelvin P Jordan.

The workshop presented in Chapter Three was planned and designed by myself, including the pre-workshop Delphi fashion survey and the workshop material including small groups 'tasks'. As I was not able to attend the workshop meeting in person, it was led on my behalf by Professor Danielle van der Windt and a panel of co-convenors who led the small group discussions, Professor Daniel Cherkin, Professor Peter Croft and Professor Maurits van Tulder.

The statistical analyses were planned and undertaken by myself with advice and guidance from my supervisor Dr Kelvin P Jordan who is a Reader in Biostatistics. The interpretation, synthesis and discussion of the findings in this thesis are my own.

Abstract

Findings from randomised controlled trials (RCTs) on primary care treatments for non-specific low back pain (NSLBP) often show modest or non-significant differences in responses to treatments. The overall response to treatment within arms, however, is often large. This raises the question of the non-specific effects associated with using the treatments and whether the size of these non-specific effects is much larger than the size of effects associated with the specific components of treatments. Non-specific effects in clinical trials, defined in this thesis as the effects on the overall improvement of symptoms (i.e. response to treatment) that is not attributed to the treatment itself, contribute to the clinical course of symptoms and can be related to the patient, the symptoms, the healthcare practitioner, the communication between the patient and practitioner, the nature of treatment provided and the setting and environment of the clinical encounter.

The objectives of this study were examining: 1) the pattern of within-arm overall responses to treatments in RCTs on non-specific low back pain; 2) sources of variation in responses to treatments by investigating the association of non-specific factors with overall responses to treatments; 3) the influence of patient characteristics on responses to treatments using individual patient data from RCTs; and 4) whether merely participating in RCTs adds to the size of response to treatments (the 'trial effect').

The findings suggest that responses to treatments for NSLBP follow a pattern of an early large improvement in symptoms within 13-27 weeks of starting treatment

followed by a smaller further improvement. This pattern was common to arms of RCTs regardless of the type of treatment. There was evidence that participants who had back pain episodes of less than 13 weeks showed larger responses to treatments than those with longer duration. There was weak inconclusive evidence for the association with age, gender, history of back pain, overall trial quality, adequacy of patient blinding and adequate compliance. There was no evidence that participating in RCTs led to larger improvement in back pain symptoms compared with participating in cohort studies. In conclusion, there is some evidence for the association of factors that are not related to the treatments with responses to treatments for NSLBP. Insufficient data hindered the assessment of other non-specific factors that were considered to be important, such as practitioner-patient communication.

Acknowledgements

There are a number of people who I would like to thank for enabling this project to materialise. First and foremost, I feel indebted to my mentor Professor Peter Croft not only for helping me capture the particular observation that made the nucleus for this work and develop it to a workable project, but more significantly for his role in nurturing my interest in research to a serious commitment and career and for his continuing guidance throughout. I would not be overstating the fact if I said I owe him my matured interest in research.

My huge thanks also go to my supervisors, Professor Danielle van der Windt and Doctor Kelvin Jordan. I have been privileged with the benefit of their experience and knowledge and the unlimited access and time commitment. All this I felt was immensely satisfying as it was coupled with the sense of humour and fun that often permeated our regular meetings and communications. I thank them for making the last four years, albeit with the not-unexpected downs and lows, a uniquely enjoyable and satisfying experience. I look forward for future opportunities to working with them again.

I would like to thank Professor Elaine Hay for her overall supervision, guidance and support throughout the project. I also would like to thank her for her role in helping me develop and advance my career as a researcher.

My thanks also go to other research members of our Centre. In particular I will not forget to thank those who helped me in the very early days as I was preparing to propose this project for the fellowship. Namely, I thank Dr Kate Dunn, Professor

Krysia Dziedzic, Professor Nadine Foster, Dr Jonathan Hill, Dr Kika Konstantinou, Dr Martyn Lewis, Professor Christian Mallen, Dr Mark Porcheret Dr Jane Richardson and Ms Gail Sowden. I also would like to thank Dr Kate Dunn again for her involvement early in developing my project idea and also for making herself available for my questions and 'shouts for help'.

I would also like to thank the Low Back Pain Forum members who participated in my workshop in the 2009 International Forum meeting in Boston USA. In particular, I would like to thank Professor Daniel Cherkin, Professor Maurits van Tulder for helping in running the workshop and eventually enabling me to conclude a successful study out of it. This has contributed to my PhD project.

And finally but most importantly, as my work on this project was on part time basis alongside my busy clinical work, it inevitably spilled into my home and family life with the expected consequent pressures and strains. My most great debt of gratitude, therefore, goes to the person without whom I can simply say I would not be able to complete this project, Ulrike, my wife. In addition to her commitment to her professional career, her unconditional and unlimited accommodation and support for me did not wane throughout. Just to give a glimpse, at the same year I started the project, Lena our daughter was two and Leo was born. Life, for all of us, was never more hectic or more enjoyable ever after!

List of abbreviations

AD	Aggregated data
CI	Confidence interval
IPD	Individual patient data
NSLBP	Non-specific low back pain
ODI	Ostwestry disability index
RCT	Randomised controlled trial
RMDQ	Rolland Morris disability questionnaire
SD	Standard deviation
SMC	Standardised mean change
SMD	Standardised mean difference
VAS	Visual analogue scale

Table of contents

List of figures

List of tables

Chapter One

Introduction and thesis layout

1.1. Introduction

Non-specific low back pain (NSLBP) is a common symptom with an estimated lifetime prevalence of up to 85% (van Tulder et al 2002, WHO 2003, Krismer & van Tulder 2006) with most people likely to experience one or more episodes during their lifetime, with the majority (around 60%) still reporting pain or disability a year after their initial consultation with a general practitioner (Hestbaek 2003). This large proportion of the affected population with a long-term pain and functional disability presents a large burden on society with large direct and indirect healthcare costs (Andersson et al 1999, Maniadakis et al 2000).

1.1.1 Non-specific low back pain (NSLBP)

Against a background of a plethora of names and diagnostic labels used for back pain (Quebec Task force on Spinal Disorders, 1987), the name 'non-specific low back pain' was gradually introduced in the eighties and commonly adopted by researchers since. Low back pain is considered to be non-specific in the majority of cases, estimated around 90-95% (Krismer & van Tulder 2006). In the remaining 5-10% cases a specific pathology is identified, such as a degenerative or inflammatory condition, disc herniation or spinal stenosis, infective, neoplastic or metabolic bone disease, trauma or a congenital disorder. The National Institute for Health and Clinical Excellence (NICE) defined NSLBP as 'tension, soreness and/or stiffness in the lower back region for which it is not possible to identify a specific cause' (Low back pain: NICE guideline 2009). This label, however, in the absent specific diagnosis, covers a heterogeneous population of patients with a wide range of characteristics (Coste et al 1991, Waddell 2005, Ben Debba et al 2000).

Although it is estimated that 90% of patients with NSLBP are expected to stop consulting healthcare professionals with back pain symptoms within three months (Croft et al 1998), evidence shows that many (28%-90%) will still be experiencing low back pain and related disability 12 months after the start of pain and many will still be on sick leave from work for back pain (Croft et al 1998, Hestbaek et al 2003, Henschke et al 2008, da Costa et al 2009). An increasing number of studies have been conducted to understand the course and outcome of NSLBP. There is growing evidence that back pain is a chronic condition that runs a persisting and fluctuating course (Dunn et al 2008, von Korff & Dunn 2008). Studies also suggest that a large proportion of patients with chronic low back pain also report widespread pain (Natvig et al 2001, Bergman et al 2001).

The back pain symptoms, the fear of them worsening and recurring impact the individual leading to loss of function, restriction in daily activities and participation in work and social activities, representing a burden as defined by the WHO International Classification of Functioning, Disability and Health (ICF) (2004). The burden on society is represented by direct costs related to healthcare service use and much larger indirect costs related to loss of productivity and work days. Maniadakis and Gray (2000) estimated that the direct costs in the UK in 1998 were £1.6 billion while the overall direct and indirect costs ranged from £6.6 billion to £12.3 billion, depending on the costing method. Similar larger proportions of indirect costs were found in Sweden and the Netherlands (Moffett et al 1995).

The current guidelines on early management of non-specific low back pain in the UK focus on the principles of promotion of self-help and keeping active, offering drug treatments (Paracetamol, NSAIDs, weak opioids and tricyclic

antidepressants) and considering exercise programmes, a course of manual therapy or a course of acupuncture (Low back pain: NICE guideline 2009).

1.1.2 Effectiveness of treatments for NSLBP and the non-specific effects

A consistent finding from many of the increasing number of randomised controlled trials (RCTs) on primary care treatments for non-specific low back pain is the modest and often non-significant differences in responses to various treatments (van Tulder et al 2006 et al, Keller et al 2007, Machado et al 2009). This is true in trials in which comparative arms include active treatments (e.g. Frost et al 2004, UK BEAM 2004, Carr et al 2005, Heymans et al 2006, Chown et al 2008), usual care (e.g. Moffett et al 1999, Hay et al 2005, Thomas et al 2006, Jellema et al 2005, Johnson et al 2007) or placebo or sham treatments (e.g. Beurskens et al 1995, Licciardone et al 2003, Rouff et al 2003, Hoiriis et al 2004, Katz et al 2005). Of equal interest is the observation of the large size of the overall within-arm response or symptom progression in these same trials (the measured change in outcome from baseline), indicating that patients on average show large improvement in clinical trials.

Although the observed difference in within-arm symptom progression in RCTs is attributed directly to the specific effect of the index trial treatment (the “treatment effect”), the overall within-arm change in symptoms will depend not only on the specific effect of the treatment. It will also represent the effects of more general factors that are not specific to the treatment used. For this reason and in keeping with the supposition in this thesis that the change in symptoms over time in trial arms is not exclusively related to the treatment used, the term ‘progression of symptoms’ will be used to refer to loosely to overall within-arm change of outcome

scores over time. As will be described in detail in the following chapter, these factors can be related to the patient (demographics, cultural background), the symptoms (severity and duration), the healthcare practitioner (skills and previous experience), the nature of treatment provided (invasiveness and intensity) and the setting and environment of the clinical encounter (Turner et al 1994, Di Blasi et al 2001, Thomas 1994, de Saintonge et al 1994, Kleijnen et al 1994, Stewart 1995, Schouten et al 2005, Smith 2009, Shaw et al 2009, Kaplan 1989, Hodges 1996). A number of names and definitions have been suggested for these factors and their effects on the course of symptoms (Di Blasi et al 2001, Katz et al 2005, Ong et al 1995, Grünbaum 1981, 1985, Moerman et al 2001, Paterson et al 2005) reflecting the complexity of this area, as will be discussed in the next chapter.

There is currently no formal definition for the non-specific effects on which researchers and clinicians agree. In this thesis the name ‘context effect’ (Di Blasi et al 2001) was adopted to refer to the change in patient symptoms that although happens when a’ treatment’ is used, is not thought to be related to the specific ingredient or component of the treatment. It is associated with the process of seeking or using the treatment. The factors that contribute to the context effect will be referred to as non-specific factors. These are the factors that are related to the patient, the symptoms, the treatment, the practitioner and setting. The effect of each particular non-specific factor, e.g. age, or symptom duration, will be called the ‘non-specific effect’. The ‘context effect’ describes the collective of all the relevant non-specific effects.

It is appropriate that rigorous attempts are made in explanatory RCTs to account for the effects of non-specific factors across the treatment arms to be able to

provide evidence for the specific effect of the treatment examined. However, the general observation of large size of symptom improvement to a wide variety of differing treatments for back pain in clinical trials with no clear superiority of most treatments raise the questions of the influence of the non-specific factors and the size of any such influence.

The observation of small or insignificant treatment effect is not unique to trials on back pain (Finnerup 2010). However, interest in understanding the role of non-specific factors in symptoms progression in clinical trials in general and specifically in pain trials has been growing (Turner 1994).This is important for a number of reasons: (1) in clinical practice, identifying and subsequently enhancing the general context effect of care could improve clinical outcome. Clear empirical evidence for the effect of non-specific factors in clinical practice is, however, yet to be provided; (2) for healthcare service provision and planning, improving the effectiveness of treatments by enhancing the influence of non-specific factors would broaden the options of available treatments and thus facilitate provision and planning based on availability and cost in the era of limited resources; (3) for research, identifying the evidence for non-specific factors, and estimating the size of their influence on the change in outcome could provide the opportunity for a better identification of the context effects associated with using treatments. This would also help inform revising the design of clinical trials to take into account the influence of non-specific factors such as identifying subgroups of patients with particular characteristics.

The aim of this study was to examine the association between non-specific factors and symptom improvements in clinical trials of non-specific low back pain (NSLBP).

1.2 Objectives

The specific objectives were:

1. To examine within-arm symptom progression in RCTs on non-specific low back pain to assess the pattern and size of symptom progression and any variation between them.

2. To examine sources of variation in symptom progression by investigating the association with non-specific factors (represented by trial level characteristics including mean age and gender of participants, duration of low back pain and aspects of trial quality) in RCTs.

3. To examine the influence of patient characteristics (age, gender, duration of pain, previous history of back pain, expectation regarding helpfulness of treatment and preference for treatment) on symptoms progression using individual patient data from RCTs.

4. To examine whether merely participating in RCTs adds to the size of symptom progression (the ‘trial effect’).

The focus in this thesis, therefore, is on the progression of symptoms representing the overall change in outcome scores over time in trial arms rather than between-arm differences or treatment effect, as within-arm progression of symptoms incorporates the effects of non-specific factors which are central to this research. However, in various areas of the thesis points related to treatment effects were raised and addressed. This is because studying within-arm symptom progression is not commonly undertaken and comparison could only be made with studies that examined treatment effects, such as the influence of trials characteristics or patients' characteristics such as age or expectation. However, the distinction between treatment effect and progression of symptoms in clinical trials was made clear throughout.

1.2.1 Identifying and prioritising non-specific factors

The definition of non-specific factors is broad in that it covers a large group of factors and hence can be complex. However, this thesis is concerned with studying factors within RCTs on back pain and for which sufficient information is available to analyse. It is important, therefore, to note that these factors might not be the only factors that influence symptom progression. To try to be as inclusive of all the important factors as possible, I conducted a study as part of a workshop at the 2009 International Low Back Pain Forum meeting, in Boston USA. The aim was to identify and obtain consensus on the non-specific factors that are most influential on progression of symptoms of NSLBP in clinical trials.

1.3 Layout of the thesis

The structure of the thesis is summarised in Figure 1. The thesis comprises nine chapters as follows (Figure 1):

Chapter One: *Introduction and thesis layout*

Chapter Two: *The non-specific factors associated with the use of treatments*

Chapter two explores the concept of the context effects associated with using treatments with a summary of a historical background.

Chapter Three: *Identifying and prioritising non-specific factors associated with treatments for NSLBP: a consensus study and a linked discussion of the literature*

Chapter three describes the outcome of a Delphi study undertaken in association with a workshop in the 2009 International Low Back Pain Forum meeting in Boston, to identify important non-specific factors that could influence symptoms progression in clinical trials on NSLBP.

Chapter Four: *The pattern of low back pain symptom progression in clinical trials: A systematic review and meta-analysis of randomised clinical trials.*

Chapter four presents a systematic review of RCTs on primary care treatments for NSLBP to examine overall symptom progression. The objective was to find out whether symptom progression in RCTs follows a common pattern similar to that anecdotally observed in a number of RCTs. To assess whether the pattern and size of change in outcome scores were independent of the treatments used, the association between symptom progression and types of treatments was studied. Treatments were classified into index, active comparator, placebo treatment, usual care or waiting list control. A second classification was according to whether the treatment was pharmacological or non-pharmacological.

Chapter Five: *The quality of randomised clinical trials of primary care treatments for non-specific low back pain*

To support the validity of the outcome of the review presented in Chapter Four, Chapter Five describes the quality of the included RCTs. The Cochrane Back Group quality assessment tool was used (van Tulder et al 2009). This tool was updated by the Cochrane Collaboration (Furlan 2009) after the quality assessment of the studies was completed. The overall trial quality, as well as individual quality criteria (e.g. adequacy of patient blinding and of concealment of allocation to treatments), were also studied as non-specific factors for their potential association with symptom progression, included in the list of non-specific factors for the meta-regression analyses in Chapter Six.

Next, the association between non-specific factors and symptom progression in clinical trials was investigated. This was achieved in three stages outlined in the next three chapters as follows:

Chapter Six: *Factors associated with low back pain symptom progression in clinical trials: a meta-regression analysis of randomised clinical trials*

In this chapter the association between non-specific factors represented by trial characteristics and symptom progression was studied. Meta-regression analyses were carried out on the aggregated data from the identified publications on RCTs. All characteristics for which sufficient information was provided were studied, including trial setting and quality, participants' age and gender and symptom duration. The associations were adjusted for types of treatments.

Chapter Seven: *Factors associated with low back pain symptom progression in clinical trials: analysis of individual patient data (IPD)*

The outcome of the second stage of analysing the association between symptom progression and non-specific factors is presented in this chapter. Some of the trial characteristics that were studied in Chapter Six were characteristics of the

participants and their symptoms and not true characteristics of the trials. Examples include age and duration of pain. The best method to study the influence of these characteristics on symptom progression is to use IPD, which was used in this chapter. The opportunity was used to study the influence of other characteristics for which some evidence exists for their influence on symptom progression and for which sufficient aggregated data were not available in the RCTs studied in Chapter Six. The focus here was on the additional characteristics highlighted during the workshop conducted at the 2009 International Low Back Pain Forum, namely expectation regarding the helpfulness of treatments, preference for treatments and history of back pain.

Chapter Eight: *The effect of participating in randomised clinical trials on patient outcome: comparing the course of back pain symptoms in RCTs and cohort studies.*

In this third and final stage of studying the association between non-specific factors and symptom progression the focus was on the "trial effect" i.e. the effect of participating in RCTs. The theory is that the intensive protocol-led care and attention provided to participants in RCTs augments the size of symptom improvement. The evidence on this is not clear and mainly derived from fields other than musculoskeletal pain. NSLBP symptom progression in RCTs included in the systematic review was compared with the course of NSLBP symptoms in cohort studies in which participants received primary care treatments or usual care. Cohort studies represent the course of symptoms without the effect of the more stringent protocol-led methods in RCTs.

Chapter Nine: *Summary and conclusions*

Chapter Nine summarises the outcome of the study, draws overall conclusions, and discusses the implications and areas for future research.

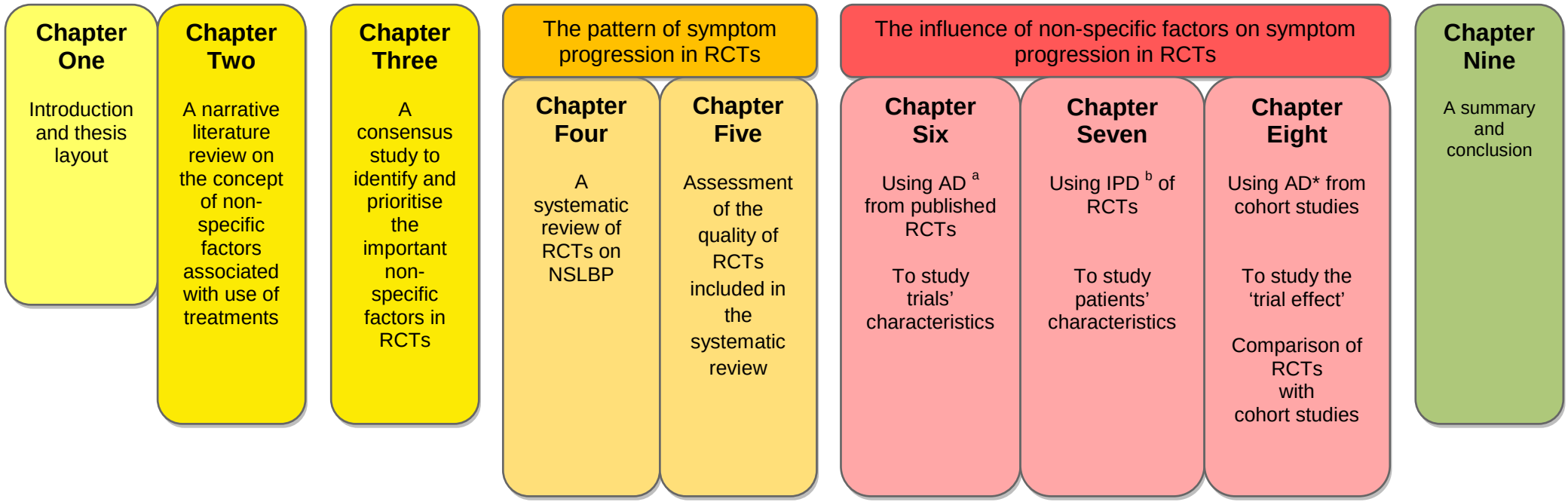


Figure 1 Outline of the thesis
[a] AD: aggregated data, [b] IPD: individual patient data

Chapter Two

The non-specific effects associated with the use of treatments

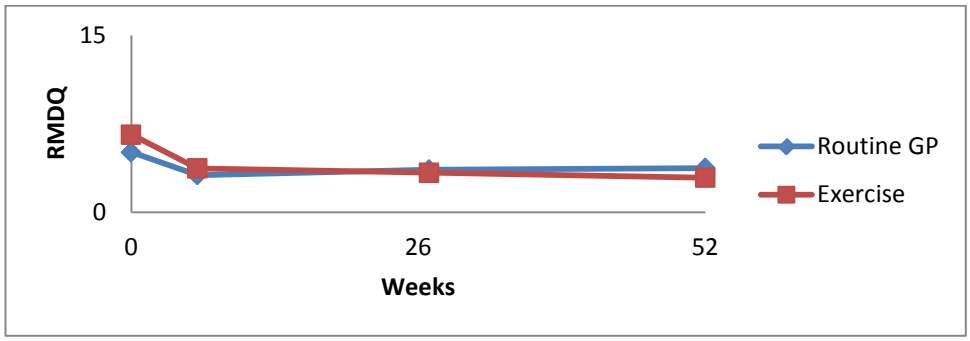
2.1 Introduction

The observation from a number of RCTs on back pain of small or insignificant treatment effects (Moffett et al 1999, Frost et al 2004, UK BEAM 2004, Hay et al 2005, Heymans et al 2006, Thomas et al 2006, Johnson et al 2007) (Figure 2.1) with ‘universal’ large overall improvement in symptoms following various types of treatments was the origin of my interest in the subject area of this project. The observation raises the questions of the role of non-specific factors associated with the use and provision of the treatments and whether the size of their effects is larger than the specific effects of the active treatments themselves that it compromises the ability to show a clear superiority of the specific effects of the treatments.

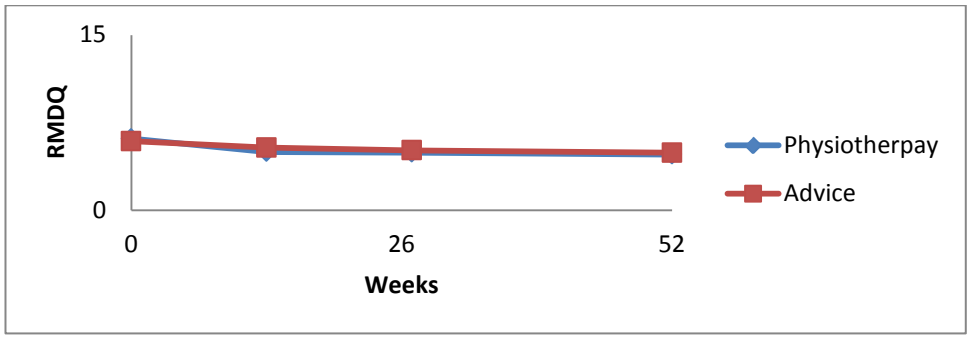
In this chapter the concept of the effects of non-specific factors associated with the use of treatments is explored with a brief review of the literature. The name the context effect will be used throughout the thesis to loosely refer to the effects of non-specific factors associated with the use of treatments but not related to the specific ingredient of the treatment.

2.2 The context effect and the specific effects of treatments

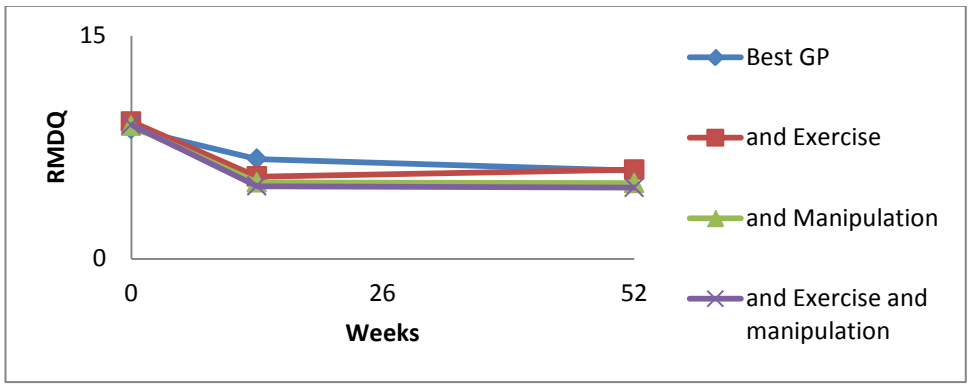
There is a well-documented history trail of the concept of ‘non-specific effects’ of treatments that have taken various names (Beecher 1955, Brokovec 1985, Gotzsche 1994, Grunbaum 1981, Grunbaum et al 1985) including the probably more commonly used term ‘placebo’ effect. The ‘incidental effects’ (Grunbaum



Moffett et al 1999

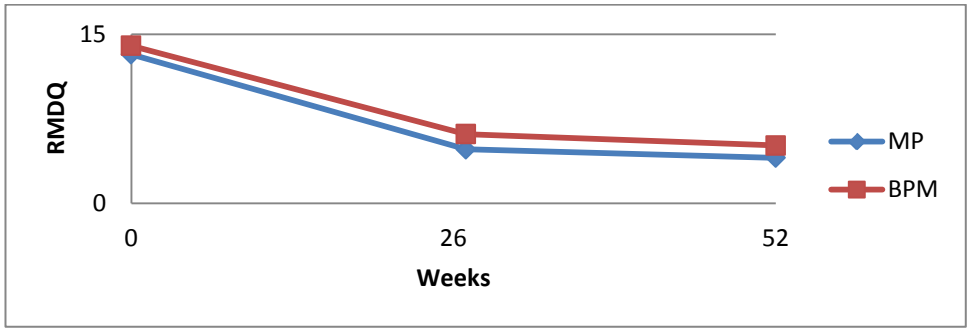


Frost et al 2004



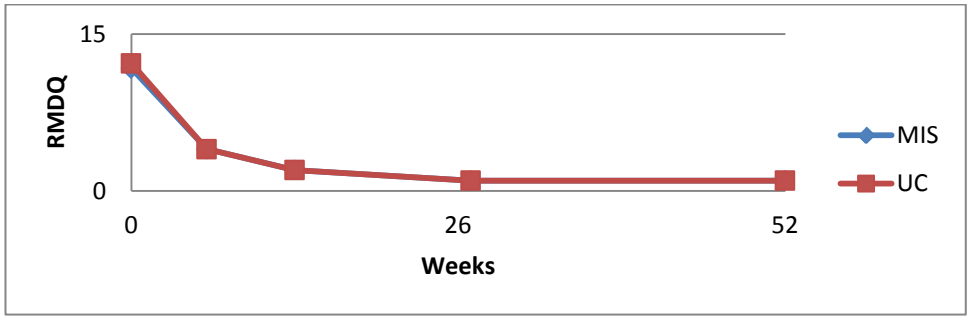
UK BEAM 2004

Figure 2.1 Within-arm symptom progression in a number of RCTs on NSLBP treatments



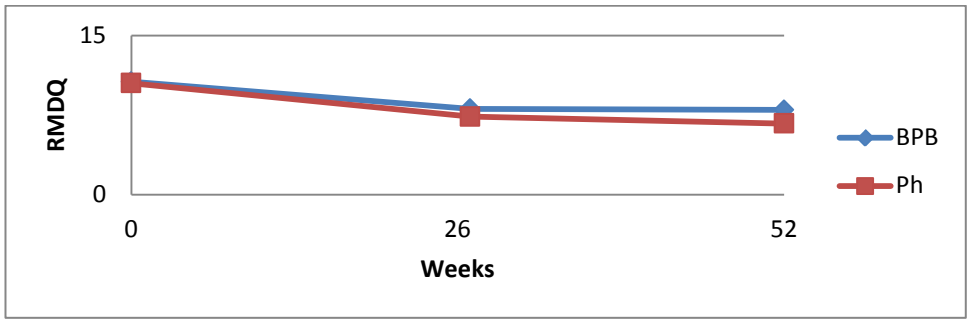
Hay et al 2005

MP: Manual physiotherapy, BPM: Brief pain management



Jellema et al 2005

MIS: Minimal intervention strategy, UC: Usual care



Johnson et al 2007

BPB: Back pain booklet & usual GP care, Ph: Physiotherapy led community based treatment program

Figure 2.1*(Continued)* Within-arm symptom progression in a number of RCTs on NSLBP treatments

1981, Grunbaum et al 1985), the ‘context effect’ (Di Blasi et al 2001) and the ‘meaning response’ (Moerman et al 2002) are some of the names suggested over time. This, however, does not represent just superficial attempts at naming a phenomenon, but in fact it nicely reflects difficulties and problems in identifying and defining this concept.

The terms specific and non-specific, for instance, did not receive universal agreement, as some found them unhelpful (Grunbaum 1981, Grunbaum et al 1985). The main problem was to what exactly the specificity is related, the treatment or the symptoms. Understanding the mechanism of action of a treatment might lead to identifying its specific effect. If this was found to help in a particular disease or symptom then this effect would be specific to that particular treatment only when used for that particular disease. If the mechanism was found to also help in another disease or symptom, then that treatment effect might not be considered specific when it is used for either of these conditions. As ‘non-specific effects’ would then be described as effects of treatment that are not expected or specified according to the mechanism of action of treatment and nature of symptoms, some preferred to describe them as ‘non-specified’ rather than ‘non-specific’ (White et al 1985). This highlights how easy it is for the discussion to become semantic and lose its core meaning.

For some treatments, it is not easy to clearly identify the specific ingredient and acupuncture is a good example. While traditionally the needling is considered as the specific ingredient, other elements of performing acupuncture such as communicating with the patient and exploring the health problem to decide the way the procedure is conducted is considered by some as important elements of

the specific ingredient of acupuncture (Paterson et al 2005). This could arguably apply to other types of treatments such as physiotherapy or chiropractic or indeed to pharmacological treatments prescribed by a practitioner.

Furthermore, even if the specific ingredient of a treatment could be easily identified, it might not be as easy to relate the change in symptoms directly and exclusively to that ingredient. This becomes particularly difficult in treatments that involve and rely on elements such as the skill of the practitioner, communicating with the patient or the setting in which it is provided. Clear and solid distinctive lines might not exist between these factors and between them and the specific ingredient of the treatment (Thomas et al 1994). Describing them as specific or non-specific, therefore, might not be universally clear or agreed. However, others do believe that these factors are distinguishable and should be addressed as such (Miller et al 2006).

It is likely that non-specific factors interact with each other and with the specific treatments and the final effect will depend on the nature of these factors and their interaction. This could therefore be the addition of the effects, an augmented overall effect or a reduction in the final effect.

Non-specific factors are sometimes called placebos and their effects described as placebo effects (Brody et al 1980, Brody et al 1985, Colloca 2004, Price et al 2001). A placebo, however, might be better reserved for describing comparator ‘treatments’ in placebo-controlled trials or clinical practice. In that narrow context, the meaning of the placebo is a control treatment with a similar appearance to the study treatment but without its essential component (Hrobjartsson & Gotzsche

2010). Early definitions of placebo and placebo effect are those of Shapiro et al (1960, 1961, 1968, 1978) who defined a placebo as “any therapy (or component of therapy) deliberately used for non-specific psychological or psychophysiological effect ... and without specific activity for the condition being treated...” The placebo effect, accordingly, was defined as “the non-specific psychological or psychophysiological effect produced by placebos”. This is an example of the earliest references to this concept, suggesting that a therapy has specific effects on a particular medical condition that are different from effects that are not specific to that therapy. As a theoretical implication of that, the magnitude of the non-specific effects in placebo controlled trials could be measured by subtracting the change in outcome scores in patients treated with a placebo treatment from that in patients in the ‘no treatment’ arm. In fact several systematic reviews have used such methods to estimate the placebo effect, for example in patients with osteoarthritis (Zahng et al 2008) or a range of musculoskeletal conditions (Hrobjartsson & Gotzche 2010, Hauser et al 2011).

For this thesis, the name context effect was adopted with a broader definition to refer to the change in patient symptoms that although happens when a’ treatment’ is used, is not thought to be related to the specific ingredient or component of the treatment. The non-specific factors could be classified into five groups (Figure 2.2): factors related to the patient, healthcare practitioner, nature and type of the treatment, the symptoms and setting or environment of the clinical encounter (de Saintonge et al 1994, Kleijnen et al 1994, Turner et al 1994).

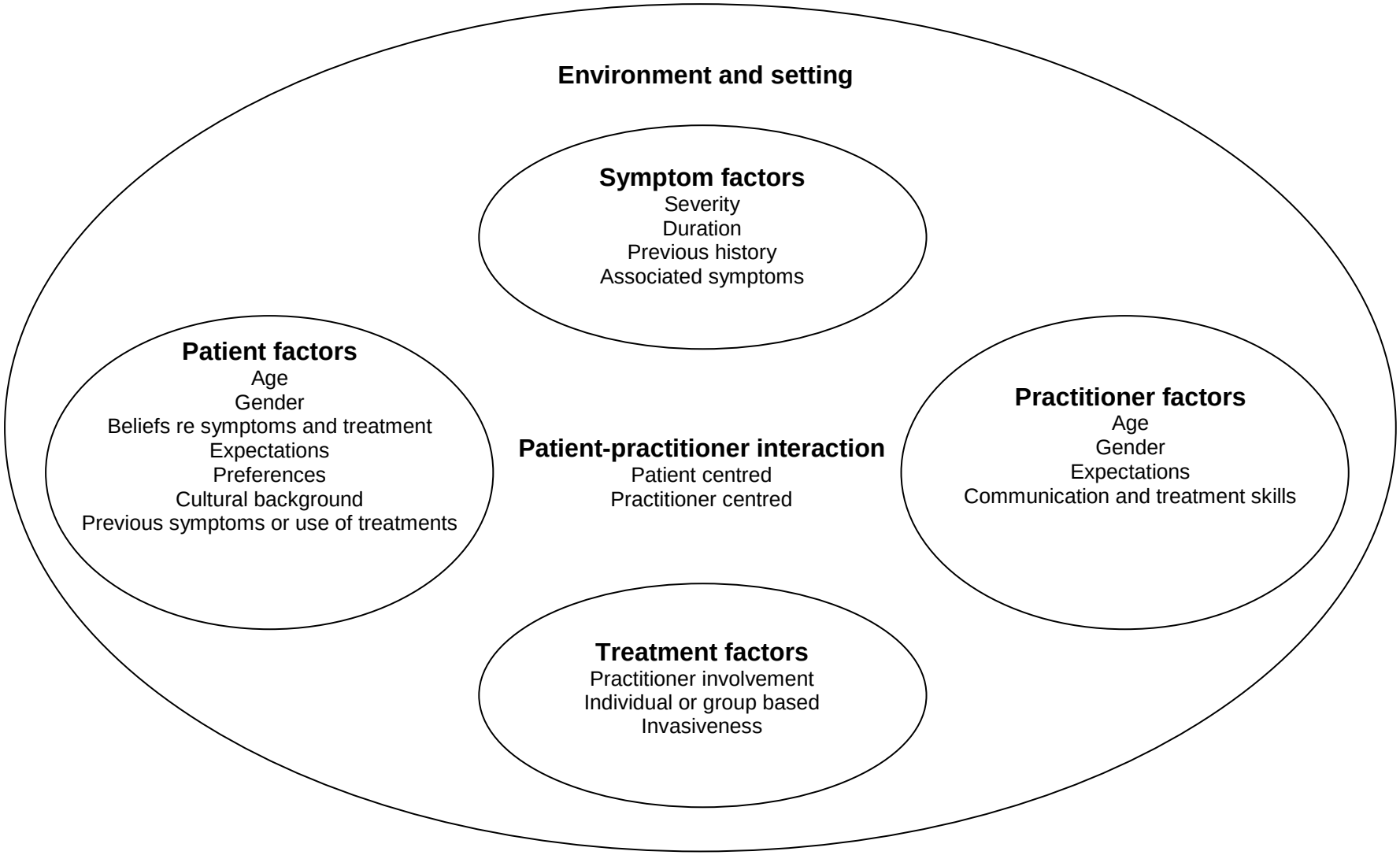


Figure 2.2 Categories of non-specific factors (including examples) associated with the use of treatments in clinical trials.

Factors related to the patients include their cultural background, demographics, beliefs, expectations and experiences with other illnesses, history of the illness or previous use of the current treatment or other treatments (Petrie et al 2005, Schouten et al 2005, 2007, Smith et al 2009). Practitioner factors include their experience with the use of the treatment, expectation and knowledge of the clinical course of the illness and their skills (Schouten et al 2005, Shaw et al 2009, Stewart et al 1995, Hodges et al 1996). Characteristics of the treatment (Branthwaite et al 1981, Colloca et al 2004) such as invasiveness, physical contact and psychological component are other non-specific factors. The environment and nature of the communication between the patient and the practitioner (Schouten et al 2005, 2007, Shaw 2009, Stewart et al 1995, Hodges et al 1996, Colloca et al 2004, Ong 1995, Gallagher et al 2005, Gorawara-Bhat et al 2007, Charles et al 1997, Coulter 1997) represent examples of other factors. These examples are by definition non-specific, i.e. not specific to the ingredient of the particular treatment used, but are associated with the process of the provision and reception of care or treatment.

2.3 Non-specific factors, natural history and clinical course

It could be argued that some factors, such as symptom severity and duration, influence the natural history of the symptoms without using a treatment. Patients' beliefs regarding their illness and symptoms, cultural background, previous experience with the symptoms, communications from relatives and friends, age or gender may also influence the progression of symptoms without using any treatment. In other words, they could be considered to contribute to the natural

history of the symptoms rather than non-specific factors. However, any influence these factors might have would still be at play when a treatment is used and they would then be considered to contribute to the clinical course of the symptoms. If the treatment was received through communicating with a healthcare practitioner in a consultation setting, then additional factors related to the practitioner and the setting will also start to have a role. In a clinical trial, additional factors related to the protocol and setting as well as other factors such as preference for a treatment and self-selection for enrolment will also play a role. Again, they will all contribute to the clinical course of the symptoms. In other words, there is no exclusivity between natural history, clinical course and the influence of non-specific factors. The focus in this thesis was to study the factors that could be associated with the course of back pain symptoms in clinical trials and to find out whether any influence these factors have is unrelated to the type of treatment, as the anecdotal observations seem to suggest.

2.4 Non-specific factors and clinical trials

The randomised placebo-controlled trial, heralded as the gold-standard method for evaluating treatment 'success', compares the effect of an 'index' treatment with the effect of another, 'comparator' treatment or 'control' condition. The efficacy of the treatment is measured by assessing the difference in the changes in pre-determined primary outcome measures between groups (Figure 2.3).

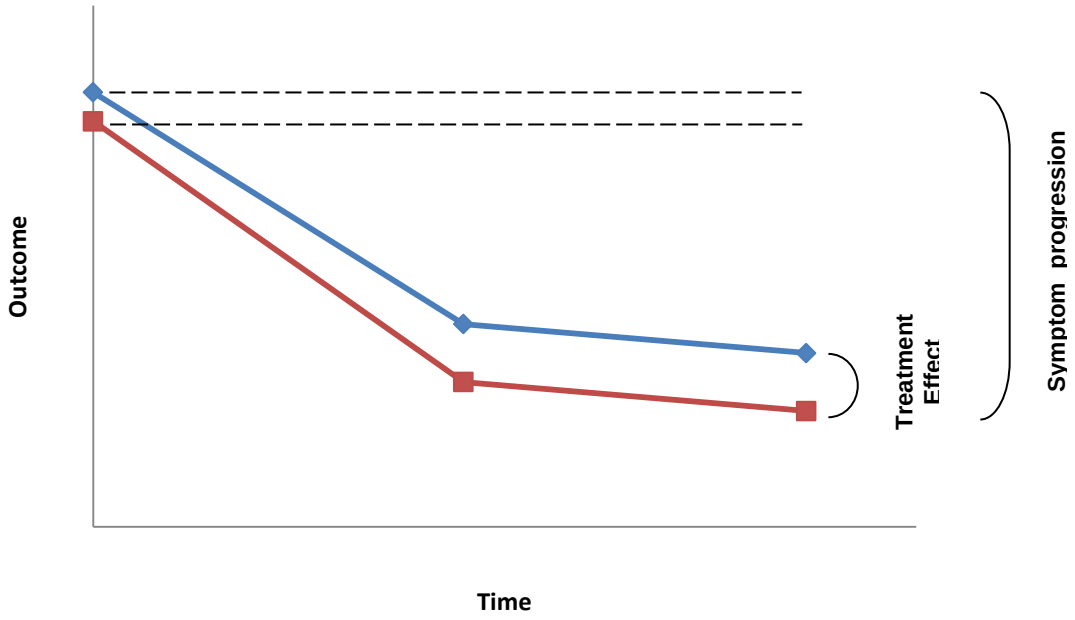


Figure 2.3 Treatment effect and symptom progression in an example of a two-arm trial.

The unique property of the RCT is that it aims to directly attribute any observed difference in outcomes to the index trial treatment. In an explanatory efficacy randomised controlled clinical trial, this difference represents the specific effect of the index treatment.

However, symptom progression in each arm will be influenced by multiple factors, in addition to the specific treatment itself (Figure 2.4). Such 'non-specific factors' can be related to the patient, the treatment, the practitioner, the symptoms or the setting of the trial. The actual participation in the trial could be considered as a factor that could result in what has been described as the 'trial effect' (Braunholtz et al 2001, Vist et al 2005).

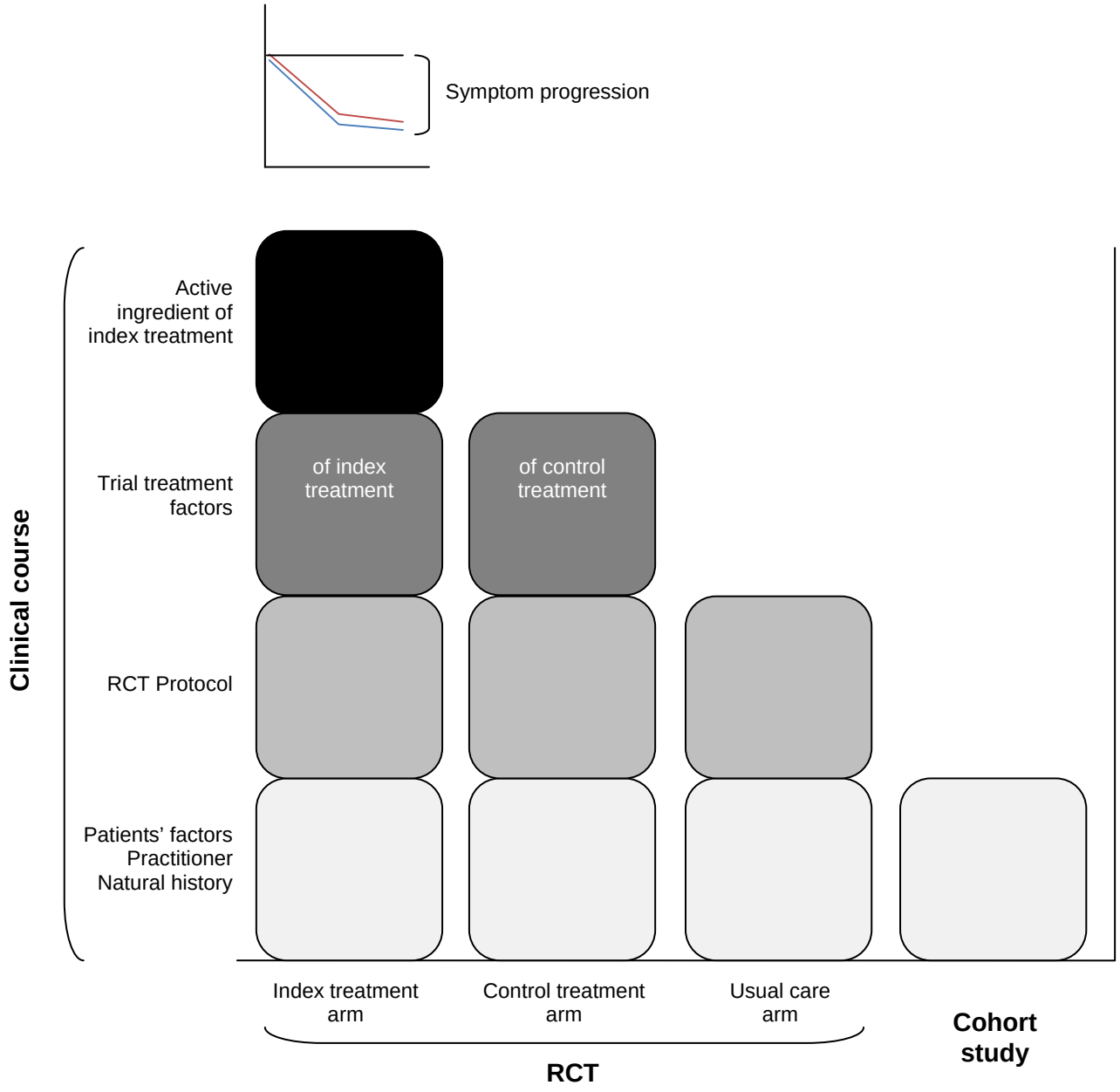


Figure 2.4 Specific and non-specific factors that could influence symptom progression in RCTs and cohort studies. The non-specific factors mentioned represent examples for illustration only. The distribution of factors is not exclusive to the places allocated for them.

These factors contribute to the symptom progression, or the course of the symptoms, in all treatment arms. If the specific effect of the trial treatment is represented by the difference between the change in outcome scores in the ‘two’ treatment arms, then the remaining size of progression of symptoms in both arms will not be related to the specific treatment effect but to the effects of non-specific factors (Figure 2.3). The size of the effects of the non-specific factors is not yet clearly estimated. The suspicion is that the size might be so large that it overshadows the specific effect of the treatments themselves.

One implication for this is that attention needs to be paid to the apparently common large change in outcome scores in all arms in these trials. Any sources of variation in the size of this change would represent the effects of factors that need to be identified to potentially aid identifying subgroups of patients with larger change in their symptoms. This fuelled the main aim of this PhD project.

Another implication is that to be able to study these non-specific factors and their influence on the course of symptom, within-arm change in outcome scores, rather than between-arm treatment effect, need to be studied. It is this within-arm symptom progression that includes the non-specific effects of all factors as well as the specific effect of the treatments and this is the reason why this study focuses on within-arm symptom progression rather than between-arm treatment effects.

2.5 Summary

It is clear that the concept of non-specific factors is complex and finding a name that describes it with all its varied and complex aspects is difficult. It was not within our aims to examine in depth the various names or to suggest a new name. The term ‘context effect’ is adopted here based on the understanding that, compared with the other available names, it represents all the factors, other than the specific ingredient of the treatment used, that would contribute to the course of back pain symptoms in clinical trials. These factors are related to the patient, the practitioner, the symptoms, types of treatment and the setting and environment. There are, of course, the caveats associated with this term as with all the other terms as outlined earlier.

Contributing to the natural history of symptoms are groups of non-specific factors related to the patient and the symptoms. The same factors will contribute to the clinical course of the symptoms when a treatment is used. The distinction between the context effect and natural history, or between the context effect and the clinical course is therefore artificial as they represent overlapping concepts with shared components.

Chapter Three

Identifying and prioritising non-specific factors associated with treatments for NSLBP: a consensus study and a linked discussion of the literature

3.1 Introduction

The apparent ‘universal’ improvement in symptoms in clinical trials following various types of treatments raises the question of the contribution of the specific versus non-specific factors associated with the use of treatments to symptom progression. A number of such factors such as patient perception about illness or treatment, patient expectation, practitioner skills and patient-practitioner interaction have been identified and their influence on benefit from treatment studied (Turner et al 1994, Crow 1999, Karjalainen et al 2004, Di Blasi et al 2001) as was discussed in Chapter Two. However, there is no empirical evidence from the literature on the hierarchy the various non-specific factors have according to their influence on the course of back pain symptoms.

To identify the non-specific factors that have influence on NSLBP symptom progression and to explore how they may rank according to the strength of that influence a two-stage design was adopted. First a Delphi study was conducted prior to a workshop at the 10th meeting of the International Forum for Primary Care Research on Low Back Pain, held in Boston, USA in 2009. The linked workshop represented the second stage in which the outcome of the Delphi study was explored to help validate the study findings. The workshop discussions were also used to further enrich and refine my understanding of the role of non-specific factors in clinical trials on NSLBP. This chapter describes the Delphi study and the workshop and provides an overview of the key findings and recommendations. The chapter concludes with a linked discussion of the literature.

3.2 Objectives

The main objectives were:

1) To identify the non-specific factors that influence NSLBP symptom progression in clinical trials, and 2) To obtain a consensus on the strength of their importance according to the strength of their influence.

A further objective was to utilise the opportunity of the workshop to discuss the current understanding of the influence of non-specific factors on symptom progression, their interaction with each other and to what extent they can be modified by treatment / targeted for intervention.

3.3 Methods

The methods involved a pre-workshop Delphi study and a workshop using nominal group discussions.

The recommended guidelines for the Delphi technique (Goodman 1987, Hasson et al 2000, Villiers et al 2005, Keeney et al 2006) were followed. These included anonymity i.e. responses that are fed back to participants are anonymised; iteration with controlled feedback i.e. using successive questionnaire or feedback to obtain successive responses; avoidance of interaction between participants; statistical summary of the group's views on the included items which is fed back to the participants; and finally the use of experts as the participants in the Delphi.

These characteristics will be described in detail below in conjunction with the methods used for this study.

3.3.1 Pre-workshop Delphi study

The study panel

Participants of a Delphi study are often referred to as the study panel. The intention in a Delphi study is not to select a random sample of participants who might have views or opinions on the subject. Rather, it is experts, defined as those who have the relevant knowledge and experience in the particular field and whose opinions are respected by their peers, are selected to participate (Goodman 1987). In addition to the experts, it is usual to also invite individuals who have specific interest in the subject area to participate in the study. The main reason for these characteristics of participants is the aim of the Delphi study, which is to arrive at a consensus or a judgement that would represent the evidence in an area where empirical evidence is lacking or difficult to establish (Goodman 1987). The application of the study could be to forecast future events based on open views of the experts, or to arrive at a consensus on a selection from a closed list of items.

For the study described here in this thesis, participants in the Forum meeting mentioned earlier were invited to take part. They were a group of clinicians and researchers with a common interest and expertise in back pain, self-selected to participate in the Forum meeting and in this workshop in particular. They, therefore, represented the experts and individuals with interest in the subject area as recommended for this type of study. The workshop and its objectives were

advertised and therefore it was assumed and expected that the participants had the expertise and/or strong interest in this subject area.

Anonymity and communication

The Delphi study described here was facilitated by the author. Questionnaires were communicated and fed back to participants throughout the study rounds with anonymity using email communication. I.e. although the participants were identifiable to the facilitator, neither they nor their responses and the information fed back to them were identifiable to individual participants.

This anonymity has an obvious advantage in that it is conducive for a truthful expression of opinions and choices. However, the downside for anonymity is that it might lead to lack of accountability, as is discussed in the limitation section below. Using this anonymous communication also ensures that direct interaction and discussion among participants before completing the questionnaires was prevented. This is important as the individual participants will provide their opinions and views without the influence of the others, although they might change or modify their view when they see the group's views in the subsequent rounds. The study described here included three rounds as described below, the third round was also used to explore modifiability and interaction of factors.

Iteration and feedback: round one

Successive questionnaires are used in a Delphi study in which information and or opinions are sought from participants which are then summarised and incorporated in the subsequent questionnaires. With the successive questionnaires fed back to the group for further opinions as they are kept informed

of the current or developing group's collective opinion. It is through this open process that offers the individual participants the opportunity to provide their true and uninfluenced opinions and views that Delphi study aims to facilitate and develop a consensus.

In the first round of the study described here, participants were invited to list all non-specific factors they considered influential on symptom progression in primary care patients with back pain. As a prompt, a list of 13 examples of non-specific factors compiled from the literature (Turner et al 1994, Thomas 1994, de Saintonge & Herxheimer 1994, Kleijnen et al 1994, Stewart 1995) (Table 3.1) was sent out to participants to build on and add additional factors. Further to this, participants were also asked to rate the strength of the influence of each factor on patient outcomes on a scale from zero to five (zero indicating no influence at all and five maximum possible influence).

Table 3.1 The initial prompt list of non-specific factors presented to participants in the first round of the pre-workshop survey

Patient factors
1. Expectations
2. Beliefs regarding the illness
3. Experience with the current treatment
4. Preference regarding the treatment
5. Cultural background
6. Duration and severity of the back symptoms

Practitioner factors
7. Experience and skills
8. Experience with the current treatment

9. Patient-practitioner interaction factors

Treatment factors
10. Treatment characteristics
11. Duration/dose
12. Whether the treatment is self-administered or involves a practitioner

13. Environment and setting factors

Other factors

Iteration and feedback: round two

Responses from round one were collected and a list of non-specific factors was compiled to include the initial prompt list and all newly added factors. This was the main aim of this round. Newly added factors that refer to a similar construct were grouped and represented by a single factor. Although participants were asked to score the factors in round one, only factors provided in the prompt list would be scored by all participants while the newly added factors would be scored only by those who provided them. The new list was then sent out to participants in round two to rate the expected strength of the influence of each factor on patient outcomes using the scale of 0-5. It is in this second round and the next when the group had the chance to score all factors, initial and newly added. The score for the factor is calculated as the mean of the scores given by the group. When two factors have a similar mean score, the factor with the largest number of highest scores was ranked higher. For instance, if factor A received four scores of (5) and factor B three scores of (5), factor A was ranked higher if their mean scores were similar.

Iteration and feedback: round three

Round two data was used to create a list of ranked factors according to the mean of scores given by participants in round two. This was then sent out to participants in this third round to score the factors for the final time and also to: 1) identify factors whose influence would depend on other variables (interaction), using free text, and 2) identify factors that could be modified to improve patient outcomes, to be selected from the provided list of factors.

3.3.2 The workshop

The workshop involved group discussions as participants were divided into three small groups (identified as Groups 1, 2 and 3), pre-selected to ensure equivalent distribution of clinicians and researchers. They were provided with the outcome of the Delphi survey represented by the ranked list of non-specific factors that was finalised from the pre-workshop survey. The groups were invited to discuss two tasks and attempt to reach consensus. The tasks were: 1) Discussing, arguing and challenging the outcome of the survey, i.e. the non-specific factors and the perceived strength of their influence on patient outcome, and 2) Identifying non-specific factors which effects could be modified within the framework of an intervention trial.

3.3.3 Organisation and identification of common themes

The workshop was my idea. I designed the pre-workshop survey questionnaires and the workshop material and facilitated the pre-workshop survey in all its rounds and analysed the results preparing them for the workshop. However, for circumstances out of my control and occurring near the time of the workshop, I was not able to attend the workshop. After communications with the meeting organisers and with the workshop co-convenors, it was agreed that the workshop would go ahead led on my behalf by one of the co-convenors, who was my supervisor. I was able to secure a live telephone connection to listen and participate in the workshop proceedings and some of the groups' discussion. The workshop and the small groups' discussions were audiotape-recorded and written notes were also obtained by the workshop leader and the groups' moderators.

Studying the notes and audiotape recordings I identified common themes regarding the most important non-specific factors, their modifiability and

interaction. The findings, themes and discussion were independently checked by the three members of the panel who led the workshop.

3.4 Results

3.4.1 Workshop participants (Study panel)

Twenty-one Low Back Pain Forum members from nine countries participated in the workshop. Participants included non-clinical researchers and practicing clinicians with active experience with, or interest in, back pain research. They included five who were practicing general medical practitioners, three rehabilitation medicine practitioners, three physiotherapists, two chiropractors, two acupuncturists, three epidemiologists and three trialists. Some participants had mixed background and interest and some were both active practitioners and researchers. All participants had particular interest, and some of them research expertise, in the non-specific effects associated with using treatments for non-specific low back pain.

3.4.2 Outcome of the Delphi study

Eighteen participants completed all three rounds of the pre-workshop survey. In the first round 50 factors (Table 3.2) were added to the prompt list of 13. Grouping new factors that represented similar constructs resulted in 17 new factors. Adding these to the prompt list of 13 resulted in a final list of 30 factors.

Table 3.2 Non-specific factors that were added to the prompt list by participants in the first round of the Delphi study

Factors added by participants in the first round	Grouping of similar factors for the second round
1. (Patient) Beliefs regarding the treatment	(Patient) Beliefs regarding the treatment
2. Meet and greet the patient 3. Clinician empathy 4. Treat the patient seriously 5. Duration of patient practitioner contact 6. Quality of relationship 7. How to greet and meet the patient 8. Clinician takes patient complaint seriously 9. Take the patient seriously 10. Quality of caring. 11. The feeling of being taken seriously 12. A nice meeting between patient and practitioner	Patient-practitioner interaction
13. Communication skills 14. Explain difficult things in an easy way 15. Explain things easy and understandable 16. Communication and ability to connect with patient 17. Explain illness and causes in a easy way 18. Ability to demonstrate confidence to patient 19. Style of language and medicine (Bio or not)	Practitioner communication skills
20. Cost of treatment 21. Perceived (by patient) cost of treatment 22. Payment for the treatment 23. Cost of treatment 24. Paying or not	Cost (true or perceived) of treatment
25. (Treatment) New or known 26. Perceived as 'high technology' (by patient and practitioner) 27. (Treatment) Traditional or technical 28. Too much treatment (Health Care)	Whether treatment is new or known, 'high-tech or traditional
29. Practitioner enthusiasm about treatment 30. 'Personality' of practitioner (enthusiastic, confident and convincing) 31. (Practitioner) Charisma	Practitioner personality and character
32. Losing dignity 33. Losing dignity	Patient losing dignity in the course of treatment
34. Self-efficacy 35. Coping style	Patient self-efficacy and coping
36. (Practitioner) Beliefs about illness and treatment	Practitioner belief regarding illness and treatment
37. (Patient) Age 38. (Patient) Age	Patient age
39. (Patient) Gender 40. (Patient) Gender	Patient gender
41. (Patient) Uncertainty	Patient uncertainty
42. (Practitioner) Expectations/trust with treatment	Practitioner expectation
43. Instrument measuring outcome	Instrument measuring outcome
44. (Patient) Experience from friends and family	Experience of patient friends and family
45. (Patient) Work / Type of work	Patient employment
46. Socialisation (Isolated vs social support network) 47 External circumstances, disability payment for example	External circumstances e.g. disability payment

48. Psychological illness - anxiety, depression	Psychological symptoms, anxiety, depression
49. Dare to let the patient decide by own experience 50. Support the patient's decisions?	Patient autonomy or freedom in making decision on management

Table 3.3 shows the final list of 30 factors ranked according to the mean scores given by participants in the study's second and third rounds. In the first round, the new factors were not scored by all study participants and therefore the scores given to them by their authors in the first round were not shown.

The list of ranked factors after the final round shows that 'Patient-practitioner interaction' is the factor considered by this group as the most influential on patient outcome. This was followed in rank by 'patients beliefs regarding illness', 'patients expectations', 'duration and severity of LBP symptoms' and 'experience and skills of practitioner'.

Table 3.3 The outcome of the Delphi study: the ranked non-specific factors and the mean group scores, on scale 0-5, according to the perceived strength of their influence on patient outcome. Factors added to the initial prompt list by participants in round one are shown in bold.

Factors in prompt list in round one	Scores*	Outcome of round two	Scores	Outcome of round three	Scores
Patient-practitioner interaction	4.22	Patient-practitioner interaction	4.58	Patient-practitioner interaction	4.32
Patients beliefs regarding illness	4.22	Patients beliefs regarding illness	4.54	Patients beliefs regarding illness	4.32
Patients expectations	3.94	Patients expectations	4.54	Patients expectations	4.29
Patient preference regarding the treatment	3.44	Patient preference regarding the treatment	4.51	Duration and severity of LBP symptoms	4.25
Patient experience with the current treatment	3.39	Duration and severity of LBP symptoms	4.48	Experience and skills of practitioner	4.20
Duration and severity of LBP symptoms	3.39	Experience and skills of practitioner	4.46	**Practitioner communication skills**	4.20
Experience and skills of practitioner	3.28	**Practitioner communication skills**	4.46	Patient preference regarding the treatment	4.19
Environment and setting	3.11	**Practitioner personality and character**	4.35	Patient experience with the current treatment	4.18
Practitioner experience with the treatment	3.11	**Practitioner expectation**	4.35	Environment and setting	4.15
Duration / dose of treatment	2.83	Patient experience with the current treatment	4.32	Practitioner experience with the treatment	4.15
Patient cultural background	2.78	**Cost (true or perceived) of treatment**	4.28	Whether the treatment is self-administered or involves a practitioner	4
Treatment characteristics	2.67	**Patient uncertainty**	4.21	Patient cultural background	3.89
Whether the treatment is self-administered or involves a practitioner	2.67	**Practitioner belief regarding illness and treatment**	3.63	**Practitioner personality and character**	3.80
		Patient self-efficacy and coping	3.60	Duration / dose of treatment	3.56
		Instrument measuring outcome	3.56	**Cost (true or perceived) of treatment**	3.50
		Experience of patient friends and family	3.50	**Whether treatment is new or known, 'high-tech' or traditional**	3.43
		Environment and setting	3.50	Treatment characteristics	3.40
		Treatment characteristics	3.21	**Patient losing dignity in the course of management**	3.35
		Practitioner experience with the treatment	3.20	**Patient self-efficacy and coping**	3.28
		Whether treatment is new or known, 'high-tech' or traditional	3.18	**Practitioner belief regarding illness and treatment**	3.20
		Duration / dose of treatment	3.14	**Patient age**	3
		Patient losing dignity in the course of management	3.14	**Patient gender**	2.80
		External circumstances e.g. disability payment	3.10	**Patient uncertainty**	2.76
		Patient cultural background	3.06	**Practitioner expectation**	2.76
		Whether the treatment is self-administered or involves a practitioner	3	**Instrument measuring outcome**	2.70
		Patient age	2.84	**Experience of patient friends and family**	2.65
		Patient gender	2.67	**Patient employment**	2.65
		Patient employment	2.67	**External circumstances e.g. disability payment**	2.63
		Psychological symptoms, anxiety, depression	2.60	**Psychological symptoms, anxiety, depression**	2.43
		Patient autonomy or freedom in making decision on management	2.10	**Patient autonomy or freedom in making decision on management**	2.31

* In round one, only factors provided in the prompt list were scored by all study participants.

3.4.3 Outcome of the workshop

Task 1: Identification of most important non-specific factors

The three groups discussed the outcome of the pre-workshop survey i.e. the ranking of the 30 factors. The outcome was agreed and confirmed that the five non-specific factors most influential on symptom progression were, in a descending order: patient-practitioner interaction, patient's beliefs regarding the illness, patient expectations, duration and severity of LBP symptoms and experience and skills of the practitioner.

Group 1 agreed that patient-practitioner interaction was most important but also considered patient-related factors (cultural background, uncertainty, decision making, health literacy) and treatment-related factors (such as costs, or self-administered care) as next in importance.

Group 2 felt that the context in which each non-specific factor operated was important as this could influence the effect on symptom progression. This group questioned whether some 'non-specific factors' such as number of treatment sessions or duration were actually specific to the particular treatment or whether their effect is generic across treatments. There was a consensus that accurate description of a factor as 'non-specific' may be difficult because it may not be easy to separate a factor from the context of its delivery.

Group 3 discussed the difficulty in assessing the importance of individual factors because of the expected interactions among factors. All groups agreed on a general view that non-specific factors are inter-related in complex ways.

Task 2: Identification of modifiable non-specific factors

The groups discussed the ways the most important non-specific factors might be modified in the context of an intervention study. One group focused on ways to improve patient-practitioner interaction. They considered the key components of the interaction to be time, empathy, listening, showing concern, creating a partnership and shared decision making. Practitioner communication skill was considered crucial and training was highlighted as a good example of a way to modify this factor. The opportunity of using patient feedback to improve communication was also discussed.

Other suggestions included public health campaigns and web-based programmes to improve health literacy and influence patient expectations and beliefs. Patient advocates, methods to improve shared decision-making, and managing the duration of the consultation were also discussed as examples of ways to modify these non-specific factors and influence patient outcome.

3.4.4 Common themes

The following are common themes identified from notes and audiotape recording of the workshop:

Theme 1: *Defining the individual non-specific factors*

A common consensus on the meaning and definition of individual non-specific factors is important. For example, it is not clear what the actual meaning of patient-practitioner interaction is and what constitutes an optimum interaction. Although it was acknowledged that standard models of communications were already available, it was not clear how generalisable those models were to various

types of interventions and situations and how their influence on symptom progression could be measured.

Theme 2: Stages of harnessing the effects of non-specific factors

Harnessing the context effect would involve three stages: identifying the important modifiable factors, designing instruments to measure their influence, and developing methods to modify their influence on patient outcomes. This is a very important aspect of studying non-specific factors. However, addressing it was considered to be beyond the time and capacity in this workshop.

Theme 3: The patient perspective

A better understanding of the non-specific factors and their relevance for patient outcome is achieved through exploring patients' views on their importance. An approach for such exploration would involve qualitative research.

Theme 4: The interaction and interdependency of non-specific factors

The majority of non-specific factors were judged to be directly or indirectly dependent on each other in their influence on symptom progression. For example, patient's preference for a particular treatment would likely interact with and be influenced by previous use of the same treatment, contact with friends or family members who had used the treatment, or having trust in a practitioner's recommendation, which itself might be influenced by various factors including the practitioner's communication skills or expertise in this particular treatment. Preference for treatment might also be related to expectation of benefit from that treatment.

Theme 5: Modifiability

Many of the listed non-specific factors are modifiable. It was suggested that ways of modifying patient-practitioner interaction could revolve around targeting its components, namely the patient, the practitioner, and the interaction between them. Examples of strategies suggested for improving practitioner communication skills included increasing the flexibility of consultation duration, learning techniques for effective reassurance of patients, and using patient feedback.

3.5 Discussion

The aim of this study was to develop a consensus on the hierarchy of the non-specific factors that influence NSLBP symptom progression according to the strength of their influence. It also aimed to utilise the opportunity of the workshop discussions to enhance the understanding of the nature of non-specific factors and their influence on NSLBP symptom progression.

A consensus was reached that patient-practitioner interaction was the most influential on NSLBP symptom progression among a total of 30 non-specific factors.

3.5.1 The Delphi study

The aim of this study was to identify the non-specific factors that could influence low back pain symptoms progression and to evaluate the comparative strength of their perceived influence on patient outcome. As the evidence from the literature was lacking on the comparative strength of their influence on the course of back

pain symptoms, the Delphi method was adopted to arrive at a consensus among experts in this area to achieve the objectives outlined in this study.

The key feature that gives the Delphi method its ability to provide strong evidence on a specific topic, when empirical evidence is absent, contradictory or insufficient, is its ‘safety in numbers’. In other words, a number of views, opinions and judgements on a specific area when combined and summarised is likely to provide stronger evidence than individual opinions and views. Before this method was developed in the 1950s (Goodman 1987), focus groups or individual experts opinions were examples of what was normally used, of which focus groups discussions were considered the most superior. However, a number of significant caveats were identified in this method which were directly addressed and overcome by the Delphi method. These include domineering personalities of some experts that might negatively impact the validity of the final conclusions or outcomes; the negative effect of hierarchy in professional status or position among the experts, where some participants might not be willing to voice opinions that would seem to undermine this; and the vocal minority who might influence the outcome against the silent majority. The Delphi method, with its anonymity and lack of direct interaction among participants and iteration and repeated collection of participants’ views address and overcome these caveats and serve to provide a valid consensus on a particular area.

There are a number of limitations, however, that are specifically related to the Delphi method and these will be outlined at the end of the chapter.

3.5.2 The outcome of the workshop in context

A number of themes identified from the workshop would merit an in-depth discussion.

3.5.2.a The most important factor: Patient-practitioner interaction

The strong consensus in this workshop that patient-practitioner interaction as the most important non-specific factor echoes a long-held belief both among health professionals and the public regarding the importance of this interaction. Developing and improving practitioner communication skills are established components of medical training. There is evidence for their influence on symptom progression, both from clinical studies (Turner et al 1994, Stewart 1995, Di Blasi et al 2001) and from experimental work such as that conducted by Colloca et al (2005). In a survey conducted by Petrie et al (2005), patients attending a pain management clinic for the first time were asked what they considered most important in the care they received in the clinic. Understanding the nature of their pain through practitioner communication was rated the highest source of patients' satisfaction with the care they received, rated as highly as pain relief itself.

The concept of 'patient-centeredness' (Stewart 1995) was developed in the eighties to counter concerns about limitations with the paternalistic and provider-centred approach prevalent in medical settings up to that time. Various socio-political changes, including development in medical laws and ethics in the 1960's and the more recent advances in public access to health information facilitated greater patient autonomy (Charles et al 1997). Healthcare professionals started to receive training in active listening and to encourage patients to express their beliefs and expectations. The concept of shared decision-making (Charles et al

1997, Katon & Kleinman 1981, Coulter 1997) to support patients to actively participate in the decision-making process about their management became increasingly popular.

The influence of ‘patient-practitioner interaction’ on symptom progression depends on variables such as practitioner communication skills and patients’ ability to engage and take part in the interaction (Schouten 1997), patient expectations and beliefs, gender and cultural and socio-economic background (Smith et al 2009, Shaw et al 2009, Schouten et al 2007). It also depends on the duration of the interaction and the environment in which it takes place (Ong et al 1995). Key elements of practitioner-patient interaction discussed in the workshop related to specific aspects of the interaction such as empathy, listening, exploring concerns, creating a partnership, and encouraging shared decision making.

Beyond identifying patient-practitioner interaction as the most important modifiable factor, it is important to have agreed methods or instruments to measure its effect. This is not the case currently, although attempts in this direction have been made to measure verbal and non-verbal elements of the interaction (Hodges et al 1996, Gallagher et al 2005, Gorawara-Bhat et al 2007). Furthermore, because of the nature of this factor, being composed of a number of components, it is important to agree on whether to measure the effects of its individual components (for example patient’s cues, practitioner cues, length of interaction) or to measure its overall effect as a total unit. Whether to use an objective or subjective measure or a combination of both is another issue that requires addressing.

3.5.2.b Patient perspective

Exploring the patient perspective was considered as an important part of a comprehensive approach to addressing the issue of the context effect. Studying the factors that patients consider as relevant and important as they seek medical care would enable us to target these factors in clinical trials.

3.5.2.c The interaction and interdependency of non-specific factors

An issue that was discussed in the workshop was whether the influence of non-specific factors was generic and fixed across all types of patients, practitioners, treatments and settings. Although factors such as expectation and patient-practitioner interaction could be studied as individual entities, their effects might vary depending on their interactions with other factors such as patients' education and literacy (Smith et al 2009), past pain experience, cultural factors, (Sanders et al 1992, Schouten et al 2007) and practitioner attitude and beliefs about the illness and treatment (Ostelo et al 2003).

An example of a non-specific factor that has been studied for dependency on other factors is patients' expectation from the treatment. Studies have suggested that expectation can be influenced by patient beliefs (Kincheloe et al 1991, MacDonald et al 1980) and by verbal suggestions (Kirsch et al 1988, Benedetti et al 1999, Pollo et al 2001, Benedetti et al 2002). This type of influence has been described as conscious conditioning (Reiss 1980, Rescorla 1988, Benedetti et al 2003) in which non-specific responses could be consciously influenced, for example, by explaining to the patient the mechanism of action of a particular treatment. The influence has been shown to be bidirectional, i.e. contrasting verbal suggestions modulated response in different directions (Benedetti et al

2003). Therefore, although it might be easier to study individual factors and try to harness their effects, the outcomes might be misleading if the interaction with other factors is not taken into account. This might be one explanation for the contradictory findings from trials that attempted to explore the effect of an individual non-specific factor such as patient expectation on treatment outcome (Jackson et al 2001, Jackson & Koenke 2001, Verbeek et al 2004, Gremeaux et al 2007, Linde et al 2007, Zens & Strumpf 2007, Foster et al 2010, Wasan et al 2010).

The influence of factors such as patient expectations, cultural background, or practitioner skills, starts at the time the patient makes the decision to seek medical or health care, i.e. before actually contacting a health practitioner or using the treatment. This represents one challenge for attempts to measure the effects of such factors on symptom progressions. However, it could arguably present an opportunity as some factors, such as expectations regarding treatment, could be 'modified' before the patient meets the practitioner. This could be done through general public health education or in a more specific way targeting patients at healthcare provision centres.

3.6 Limitations

The outcome and conclusions arrived at through this work have a number of limitations. Some of these are recognised limitations of the Delphi method itself. The first is the anonymity of the process, as it is suggested that although this is expected to be a strength of this method, it could also lead to a lack of

accountability, as participants might provide incorrect or baseless views without the ability to be directly challenged (Goodman 1987, Hasson et al 2000). However, participants are usually selected, or self-selected, for their interest or expertise in the area and their willingness to participate. The repeated process of providing opinions on the issues in the survey would also allow the group to judge the information provided. The workshop meeting in this study also provided further opportunity to discuss, argue and challenge the outcome of the survey and any associated assumptions to further validate the conclusions, without impacting the particular outcome of the survey. The survey conclusions and outcomes were confirmed in the workshop discussions, which strengthened the validity for those outcomes.

Another limitation of the Delphi method is its reliability, i.e. the extent to which it produces similar results on another occasion among another group of participants. In fact, there is no evidence of the reliability of the Delphi process (Hasson et al 2000). Obviously participants in this survey were not all those who have expertise or interest in back pain research. It is possible, therefore, that another group of experts, or indeed including all experts in this field might indeed provide a different outcome. Formal qualitative methods have been suggested to address this problem (Hasson et al 2000). However, this study represented an opportunity for discussion and not as solid undisputable evidence and therefore formal qualitative methods were not used.

Participants of the workshop were not purposefully selected in a comprehensive or exhaustive way. Rather, participants were self-selected for this workshop. However, the sample included researchers and clinicians from various disciplines,

countries, and health care settings, representing a wide range of views and perspectives.

Finally, the time available was too limited to discuss all aspects of this complex subject, although the pre-workshop survey helped prepare participants for the meeting and introduce them to the various concepts. Important consensus points were reached and relevant issues were raised that could be developed in further research.

3.7 Summary

The intention of this study was to reach a provisional agreement on the non-specific factors that are influential on responses to treatments for NSLBP. The opportunity of group discussion was also utilised to address the influence of these factors in more depth. The following conclusions provide a practical foundation upon which further investigations into our understanding of the concept could be developed:

1. Of the various factors that could influence patients' responses to treatments for non-specific low back pain, patient-practitioner interaction was identified as the most important. It encapsulates the practitioner, patient, symptoms and setting related factors as well as the interaction process and context in which they interact.

2. The need to identify important non-specific factors and their potential influence on symptom progressions from the patient point of view was recognised as an important issue to explore.

3. Agreeing on methods for measuring the context effect associated with the use of primary care treatments for back pain will be key for investigating their influence on patient outcome in clinical intervention studies.

4. The outcome of this study will inform the approach in this thesis with a focus on the most influential factors that were identified here. However, as the focus in the thesis is on factors in clinical trials, studying factors would be restrained and limited by the availability of data on such factors from clinical trials.

Workshop participants

Carlo Ammendolia, associate scientist and clinical epidemiologist at the institute of Work and Health, Jeffrey Borkan, professor and chair of the Department of Family Medicine, Warren Alpert medical school, Brown University and president of the Association of Departments of Family Medicine Memorial Hospital of Rhode Island, Tim Carey, director of the Cecil G. Sheps Centre for Health Services Research, University of North Carolina, USA, Daniel Cherkin, Senior Scientific Investigator, Centre for Health Studies, Group Health Research Institute, and Affiliate Professor, Departments of Family Medicine and Health Services, University of Washington, USA, Patrick Cote, Senior Investigator at the Centre of Research Expertise in Improved Disability Outcomes, University Health Network Rehabilitation Solutions, Toronto Western Hospital, Canada, Peter Croft, Professor of Primary Care Epidemiology, Keele University, UK, Robert Drake, Professor of Psychiatry and of Community and Family Medicine, New Hampshire-Dartmouth Psychiatric Research Centre, USA, Simon French, Senior Researcher at the Australasian Cochrane Centre and Monash Institute of Health Services Research, Maria Gram, Senior researcher from the Hospital for Rehabilitation, Oslo University Hospital, Norway, Silje Heiszter, Senior Researcher from the Hospital for Rehabilitation, Oslo University Hospital, Norway, Invild Lie Indergaard, Senior Researcher from the Hospital of Rehabilitation, Oslo University Hospital, Norway, Francisco M Kovacs, Director of Scientific Department, Kovacs Foundation, Spain, Tonny Kvaerne, Senior researcher from the Hospital for rehabilitation, Norway, Henrik Lauridsen, Director of studies, Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Martyn Lewis, Senior Lecturer in Statistics, Keele University, UK, Liv Magnussen A senior researcher at the department of Public Health and Primary Health Care, University of Bergen, Norway, David Newell, Senior Lecturer in Biomedicine and Research at the Anglo-European College of Chiropractic, UK, Glenn Pransky, Director,Center for Disability Research, Liberty Mutual Research Institute for Safety, USA, Prandeep Suri, Instructor, Department of Physical Medicine and Rehabilitation, Harvard Medical School and Researcher, Division of Research, Department of Orthopaedics, New England Baptist Hospital, USA, Maurits van Tulder Professor of Health Technology Assessment, Department of Health Sciences & EMGO Institute for Health and Care Research, The Netherlands, Danielle van der Windt, Professor in Primary Care Epidemiology, Keele University, UK.

Chapter Four

The pattern of responses to primary care treatments for low back pain:

A systematic review and meta-analysis of randomised clinical trials

4.1 Introduction

Interest in evaluating the effectiveness of primary care treatments for back pain has grown substantially over the past two decades represented by the increasing number of publications of clinical trials (Koes et al 2005, van Tulder et al 2006, Keller et al 2007). The general finding from a large number of primary care trials on back pain is a significant improvement of symptoms in all treatment arms with modest differences between them (Licciardone et al 2005, Carr et al 2005, Frost et al 2004, UK BEAM team 2004, Hay et al 2005, Jellema et al 2005, Johnson et al 2007, Leboeuf et al 2005, Burton et al 1995). Responses to treatments in a number of these trials seem to follow a pattern of improvement occurring early after the end of treatment and apparently unrelated to the type of treatment. The UK BEAM trial (2004) compared 'best GP care' with exercises and spinal manipulation, and although small differences between groups were reported, their clinical significance remained unclear. In patients with acute back pain, Hay et al (2005) found no difference in outcomes between those who received a 'brief pain-management programme' and those who received a physiotherapy treatment focussed on biomechanical dysfunction of the spine. A trial from the Netherlands (Jellema et al 2005) compared 'usual care' with a 'minimal intervention strategy' delivered by GPs intended to address psychosocial factors, and again found no significant difference between the groups. A fourth trial published in 2007 (Johnson et al 2007) compared a program of eight 2-hour group exercise session over six weeks comprising active exercise and education delivered by physiotherapists using a CBT approach in the intervention arm with a back pain educational booklet and concluded that the intervention showed small and insignificant effect at reducing pain and disability.

In these example trials, the overall changes in outcomes were large in all treatment arms and seem to follow a common pattern of rapid and large improvement in symptoms over the follow up period. Further to these anecdotal observations, I am not aware of evidence that this pattern of symptom improvement is replicated in a large number of trials across varied types of treatments for low back pain. The aim of this study was to determine whether such evidence exists.

4.2 Objectives

The objective in this study was to assess overall responses to treatments among non-specific low back pain patients in randomised clinical trials to find out whether they follow a common pattern following a wide range of primary care treatments. This was done through conducting a systematic review and meta-analysis of overall progression in pain and functional disability over time in published trials on NSLBP.

4.3 Methods

4.3.1 Criteria for inclusion of trials

4.3.1.a Trials design

Included were randomised clinical trials published in English and in which primary care treatments for NSLBP were evaluated.

4.3.1.b Participants and symptoms

Included were trials among adult patients with NSLBP which was defined as pain in the area below the lower ribs and above the gluteal folds, with no known underlying pathology (Krismer et al 2007).

Excluded were trials conducted among patients with specific LBP of an identifiable cause (e.g. cancer or arthritis), post-operative or post-traumatic back pain, or back pain associated with pregnancy or labour. Also excluded were trials conducted among healthy participants.

4.3.1.c Treatments and setting

Primary care treatments were defined as treatments that would be within the expertise of primary healthcare practitioners and would be provided or performed within the usual facilities of primary healthcare, its equivalents or departments associated with it such as physiotherapy departments, rehabilitation units and occupational healthcare departments. However, given the differences in the organisation of care between countries, the emphasis was more on the type of treatment rather than the setting and trials evaluating primary care interventions among, for example, hospital outpatient attendees or the general population were also included. Examples of treatments include Yoga, non-steroidal anti-inflammatory tablets, exercises, chiropractic adjustment, physical therapy, transcoetaneous electrical nerve stimulation (TENS), acupuncture and osteopathic manipulation. Examples of primary care practitioners include general practitioners, family doctors, physiotherapists, occupational therapists, acupuncturists, osteopaths and psychologists.

4.3.1.d Outcome measures

The initial aim was to collect data on as wide a range of outcomes as possible, to allow combining the results from multiple trials. However, outcome measures were chosen that would be used by the largest number of trials, namely the pain Visual Analogue Scale (VAS) 0-10 or 0-100, pain Numeric Rating Scale (NRS) (Guyatt et al 1987, Dworkin et al 2005, Ferreira-Valente et al 2011) and Roland Morris Disability Questionnaire (RMDQ) (24-point) (Roland & Morris 1983) or its modified versions or Oswestry Disability Index (ODI) (0-100) (Fairbank et al 1980) for back pain related functional disability.

4.3.2 Data source

The Cochrane Register of Controlled Trials (CENTRAL), the first quarter issue of 2007 accessed in April, was searched using the term ‘low back pain’ and using the MeSH Tree for the term ‘low back pain’.

As the aim was to investigate the pattern of responses to a wide range of treatments for NSLBP rather than to estimate the effectiveness of a particular treatment, an exhaustive inclusion of all trials on back pain treatments was not required. The aim was to have a large and representative pool of clinical trials that would vary sufficiently with respect to the types of treatments to achieve the objectives in this review and the CENTRAL database satisfied this aim.

CENTRAL database includes citations of published articles taken through systematic searches from bibliographic databases (notably MEDLINE and EMBASE) and other published and unpublished sources (Dickersin et al 2002). In addition, each Cochrane Review Group maintains and updates a collection of

controlled trials relevant to its own area of interest and these are called ‘Specialized Registers’. Each group may also collect items that are relevant to its own field of interest and these are known as ‘hand search results’. The registers and hand search results are assembled and collated before incorporation into *The Cochrane Library*. Each quarter, the CENTRAL dataset is re-built, using records from the four sources mentioned above, in the following order: MEDLINE, EMBASE, hand-search results and Specialised Registers. Therefore, for example, if a Specialised Register record matches an existing MEDLINE or EMBASE record, the MEDLINE or EMBASE record will be preferentially published (The Cochrane Collaboration website).

CENTRAL database was created as a repository for all citations of reports of trials identified by the Cochrane Collaboration. Due to the required quick turn-around time and relative lack of quality control, CENTRAL inevitably contains some typographical errors, duplicates, and reports of non-trials.

4.3.3 Selection of trials

The author and a reviewer (DvdW) piloted applying the inclusion criteria on a sample of 10 abstracts. The process of selecting trials for inclusion was then rolled out and independently conducted by the author and two reviewers (DvdW and KPJ) for the rest of the identified abstracts. Any disagreement on inclusion was resolved through discussion and consensus and a discussion with the third reviewer when necessary. When the information in the abstract was not sufficient to make a decision on inclusion or the abstract was not available, the full text of the trial was retrieved. The processes of reviewing the full text of the trials to

finally deciding on their inclusion were conducted by the author and a second reviewer (DvdW). For excluded trials, the reasons for exclusion were noted.

The quality assessment criteria checklist of the Cochrane Back Review Group (van Tulder et al 2003) was used to assess the quality of the selected trials. The methods used in quality assessment and the results will be described in detail in Chapter Five. The quality of the trials was not used as an inclusion criterion. The influence of quality on responses to treatments in clinical trials is investigated in Chapter Six.

4.3.4 Data extraction

The author extracted all information into a standardised form (Appendix 4.1) and a second reviewer (DvdW) who checked all extracted data. Any disagreement was resolved through consensus.

The extracted data included information to fulfil the objective relating to examining overall within-arm responses to treatments. This included:

1. Type of outcome measure: pain and/or functional disability.
2. Mean change and/or absolute scores on pain and/or disability in each trial arm at each time point, along with the standard deviation (SD).
3. Types of treatments: index, active comparator, placebo, usual care or waiting list and pharmacological or non-pharmacological treatments.

4.3.5 Data analysis

Differing units of outcome measures provided in the included trials were unified to enable comparison and converted into mean and SD. Data presented as ‘median’,

‘average’ and ‘least square means’ (LS means) were treated as estimates of mean scores (Hozo et al 2005). When SE was reported, the standard deviation was calculated as (Hozo et al 2005):

$$\sigma = SE * \sqrt{n}$$

Where σ is the observed standard deviation, *n* is sample size.

When IQR was reported, based on the properties of the normal distribution, SD was estimated as (Hozo et al 2005):

$$\sigma = IQR / 1.349$$

When range was reported, SD was estimated by dividing the range by four if sample size was smaller than 70, and by six if sample size was larger than 70 (Hozo et al 2005). When information was not available to calculate baseline SD in the study, it was calculated using the mean of SDs from comparable trials (Furukawa et al 2006, Higgins et al, Cochrane handbook, Cochrane Collaboration website)

To allow comparison between trials, commonly used follow-up times of 13, 27 and 52 weeks were selected. Other follow-up times were matched to these if they were within three weeks of the nearest selected times. Data from follow-up times that fell outside these limits were not used.

4.3.5.a The pattern of symptom progression

Examining the pattern of responses to treatments was made in three steps including a descriptive assessment of pattern, assessment of variation in sizes of responses and summarizing the overall pattern of responses, as follows:

4.3.5.a.1 Describing the general pattern of responses to treatments

Exploration of the general pattern of response was first done through visual assessment of outcome scores of pain and functional disability. These were plotted as graphic lines, using Excel 2000 (Microsoft, Redmond, WA, USA), representing baseline and follow-up mean scores for pain intensity (0-100), RMDQ and ODI at each time point for each arm of each trial.

4.3.5.a.2 Assessing variation in size of symptom progressions

To examine variation in the size of responses to treatments (heterogeneity) across trial arms, changes in outcome scores were analysed by calculating the standardised mean change (SMC) which is similar to the standardised change effect size (Morris 2000) or standardised mean difference (SMD) used to estimate the size of treatment effects. In this thesis, SMC refers to the change in outcome scores within each arm rather than between arms, as is the case in SMD. The SMC is a method of standardizing the measurement of change over time so that studies using slightly different scales, but measuring the same underlying construct, can be combined and more easily compared. This meant that by using the SMC, studies which used the modified versions of the RMDQ could be included in the same analysis as those which used the original version. Similarly, studies which used a VAS for pain could be combined into the same analysis as those using an NRS. For each trial arm, SMC was calculated by subtracting the follow-up mean score of the outcome measure from the baseline mean score and dividing by the SD at baseline. This assumes that pre-test and post-test scores were normally distributed with separate means but equal variance (Morris 2000). SMC was calculated separately for each of the three outcome measures (RMDQ, ODI and pain severity 0-100).

To compute the 95% Confidence Intervals for response sizes (SMCs), the variance (squared standard deviation, σ^2) of response size was calculated using the following formula (Morris 2000):

$$\sigma^2 = 2(1-\rho)/n\ [(n-1)/(n-3)]\ [1+n/2(1-\rho)\delta^2] - \delta^2/[c\ (n-1)]^2$$

Where: c (n-1) approximates 1 - [3 / 4(n-1) –1], ρ is the population correlation between baseline and follow-up scores which was estimated as 0.5, n is sample size and δ is the SMC. The standard error was then calculated from the SD using the equation mentioned earlier.

Participants from the same trial are likely to have similar characteristics which might lead to a potential cluster effect on response in arms of the same trial. This is particularly relevant for this study as it might contribute to any common pattern of responses. To overcome this possibility a sensitivity analysis was carried out to compare the results across all arms with the results based on one arm randomly selected from each trial.

The overall within-arm responses to treatments, plotted in forest plots representing the point estimate of response size and 95% confidence interval, was investigated for heterogeneity by computing I^2 (Higgins et al 2003). This homogeneity statistic I^2 is an approach that quantifies the extent of heterogeneity, providing a measure of the degree of inconsistency in the trial arms' results. The resulting I^2 quantity describes the percentage of total variation across trial arms that is due to between-trial arm variation. A zero% value indicates no variation at all and 100% value indicates that all variation is the result of variation between arms (rather than within- arm (sample) variation). It is calculated as:

$$\tau^2 / (\tau^2 + \sigma^2)$$

where τ^2 (Tau) is between-arm variation and σ^2 within-arm variation.

4.3.5.a.3 Summarising the overall pattern of responses

Where heterogeneity existed, assessed through the I^2 statistic, pooled estimates of responses were calculated using a random effects model weighted by inverse variance (Der Simonian & Laird 1986, Berard et al 1998). Larger trials, which have smaller standard errors, were therefore given larger weight than smaller trials. If the heterogeneity in responses to treatments was small, a fixed effects model was used.

To explain the concepts of the fixed and random effects models, variation in treatment effects (or in responses to treatments as in this review) is usually understood to have two main components: within-trial variance and between-trial variance. Analyses that assume that between-trial variance is zero correspond to 'fixed effect' model, i.e. the true symptom progressions is assumed to be the same value, or *fixed*, in each trial. It is generally accepted, however, that heterogeneity will exist between trials reflecting variation in setting, population or methods, leading to the assumption that a 'random effects' model will better represent true responses to treatment if the available estimates show considerable variation (Der Simonian & Laird 1986). In this model responses to treatments for each trial are assumed to vary around some overall mean response and that response sizes are assumed to have a normal distribution. The same concept applies in this review except that the concept concerns variation between trial arms rather than trials.

Next, the evidence was explored for the hypothesis that responses follow a common pattern regardless of the type of treatment. Analyses were therefore repeated for trial arms stratified according to the type of treatment based on whether the treatment was an index treatment, active comparator treatment, usual care, waiting list or placebo treatment. This was to assess whether the pattern of responses is dependent on whether the treatment was 'active' or 'inactive', a 'new' trial treatment or a usual care treatment. A second classification into pharmacological or non-pharmacological treatment was also used as an alternative exploratory approach. Trials on non-pharmacological treatments may show considerable differences in design, expectations regarding treatments, practitioner skills and time spent as well as other potentially important non-specific factors which may influence responses to these treatments compared with non-pharmacological treatments. Also, non-pharmacological treatments constitute the largest proportion of the treatments used for back pain.

Meta-analysis was performed using STATA/IC 10.0 software (Stata Corp., College Station, TX, USA).

4.4 Results

4.4.1 Selection of trials (Figure 4.1)

The search of the CENTRAL database yielded a total of 772 trial citations. Based on citations and abstracts, available information was sufficient to exclude 523. It was not possible at this stage to make a decision on 249 citations, either because the available information was not sufficient or no abstract was available.

Their full texts were therefore retrieved, reviewing which resulted in excluding a further 123 and including the remaining 126 publications. Eight of these publications were each a second report of the same trial and therefore information was available from 118 trials.

The most common reason for exclusion (for 203 citations) was not being a randomised clinical trial. These included reports of observational studies, methodology and protocol reports, reviews, clinimetric and biomechanics studies, studies to assess adherence to clinical guidelines, qualitative studies, validity or reliability studies, diagnostic studies, cost analysis studies, prevention studies and also editorials, letters and comments. One hundred and seventy two trials were on treatments that were not primary care treatments, the most common was surgical treatment. In 166 trials back pain was not non-specific, the most common example was back pain associated with pregnancy and labour, but this group also included back pain associated with known diagnoses such as ankylosing spondylitis, intervertebral disc herniation and malignancy. One hundred and seven trials were excluded because the populations were not from primary care or its equivalent. The final group of trials that were excluded was 89 trials that used outcome measures other than those selected for this review. Although these are single reasons to exclude trials, a number of publications had more than one reason to be excluded.

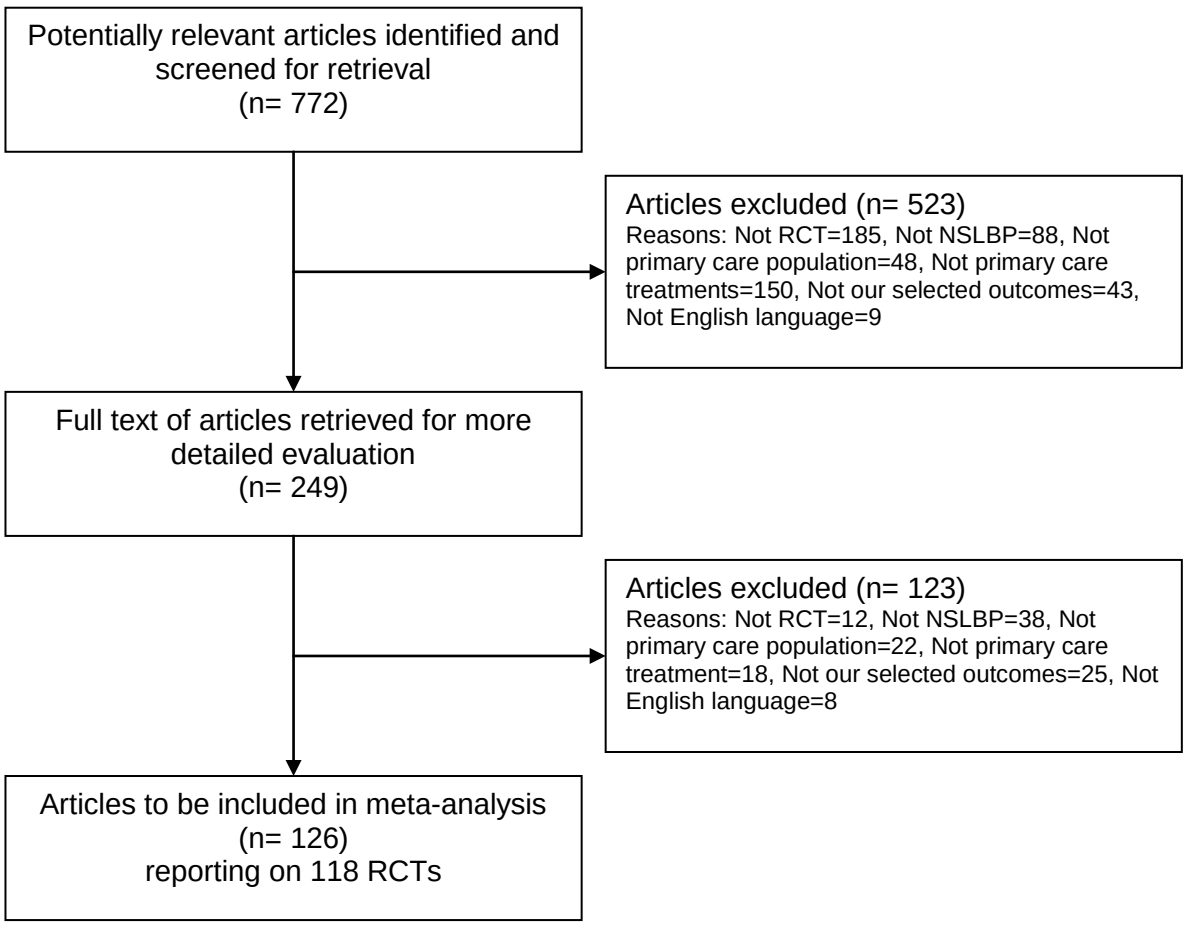


Figure 4.1 Flow diagram for inclusion of RCTs for the review
*When two articles were published for the same trial, information was extracted from both and considered as one trial. RCTs: Randomised clinical trials, NSLBP: Non-specific low back pain

4.4.2 Characteristics of included trials

Details regarding setting, populations and treatments for included trials are presented in Appendix 4.2. Trial sample size ranged from 20 to 1334 (mean 174, SD 175, median 113). The number of arms was two in 74 trials (63%), three in 31 trials (26%), four in 12 trials (10%) and five in one trial (1%). Sample size in each trial arm ranged from four to 340 (mean 67, SD 54, median 51). Duration of follow-up ranged from five days to three years. Participants' age ranged from a trial mean of 27 to 79 years (mean across trials of 43 years, median across trials of 42 years). The percentage of females in trials was 0-100% (median of 55%).

Primary healthcare settings included general practice (29 trials), occupational health care departments (20 trials) and physiotherapy departments (10 trials). Fourteen trials were conducted among the general population, 31 in mixed settings and in 14 trials the setting could not be clearly identified.

The majority of trials included in this review used another active treatment as the comparator (81 trials, 69%) and in 27 (33%) of these trials more than one active comparator treatment was used. Participants were allocated to placebo or sham treatment in 36 trials (31%) and waiting list in 11 trials (9%). Ninety-one trials were conducted to assess non-pharmacological treatments, 20 trials pharmacological treatments, five trials mixed treatments and it was not possible to classify two trials according to the type of treatment used.

4.4.3 The pattern of symptom progression

Information for VAS pain intensity was available for the follow up points of 13, 27 & 52 weeks from 104 arms in 44 trials, for RMDQ from 82 arms in 35 trials and for

ODI from 61 arms in 26 trials. Of the included trials, 30 trials were shorter than 10 weeks and therefore their data could not be included in assessing the pattern of symptom progression. However, the pattern of symptom progression in these short trials was examined and will be commented on and all these trials were included in the quality assessment.

Results were reported as mean scores and standard deviations (SDs) by the majority of trials and were used when available. Twenty -five trials reported their results in other formats such as “average”, median, range, percentage, 95% confidence interval (95% CI), interquartile range (IQR), standard error (SE) and least square means (LS means) and data were converted to estimates of mean and SD as described in Methods.

Graphic representations of overall responses to treatments are shown in Figure 4.2a, b, c representing pre and post-treatment mean scores for the three main outcome measures for each trial arm. Response lines for all three outcome measures followed a common pattern of improvement in symptoms in most arms represented by a rapid early and large reduction in mean outcome scores within the first 13 weeks followed by a slower reduction thereafter. This common pattern in responses remained when responses from only one arm were randomly selected from each trial (Appendix 4.3). This pattern appeared to be similar regardless of the type of treatment.

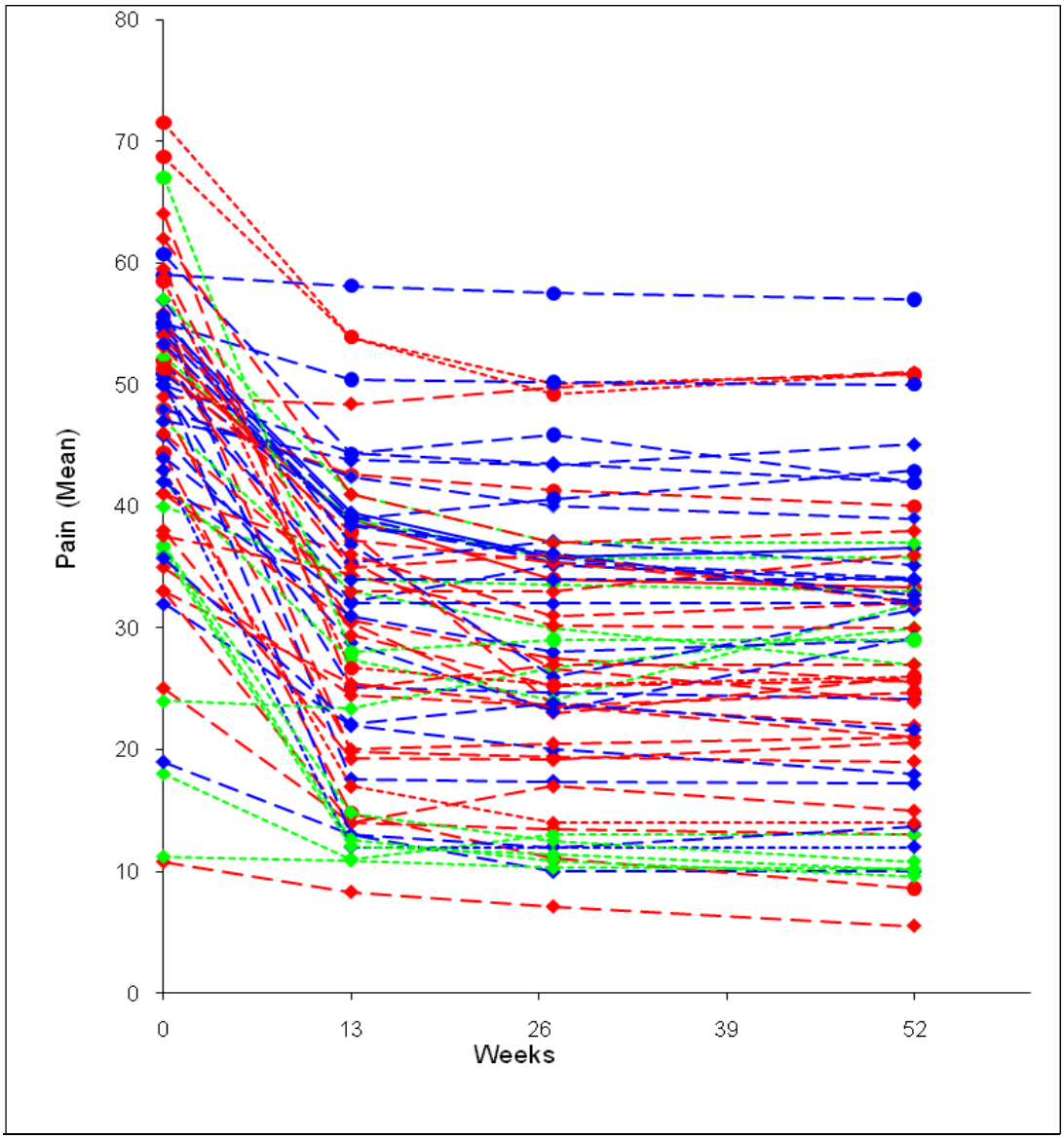


Figure 4.2a Change in pain intensity outcome scores (VAS for pain intensity) up to 52 week follow up in each treatment arm of included trials.
Each line represents a response line within each trial arm.
Red: Index treatment arm, Blue: Active treatment arm, Green: Usual care/waiting list/placebo arms.
____: Pharmacological treatment, - - - -: non-pharmacological treatment,
…….: Mixed/other.

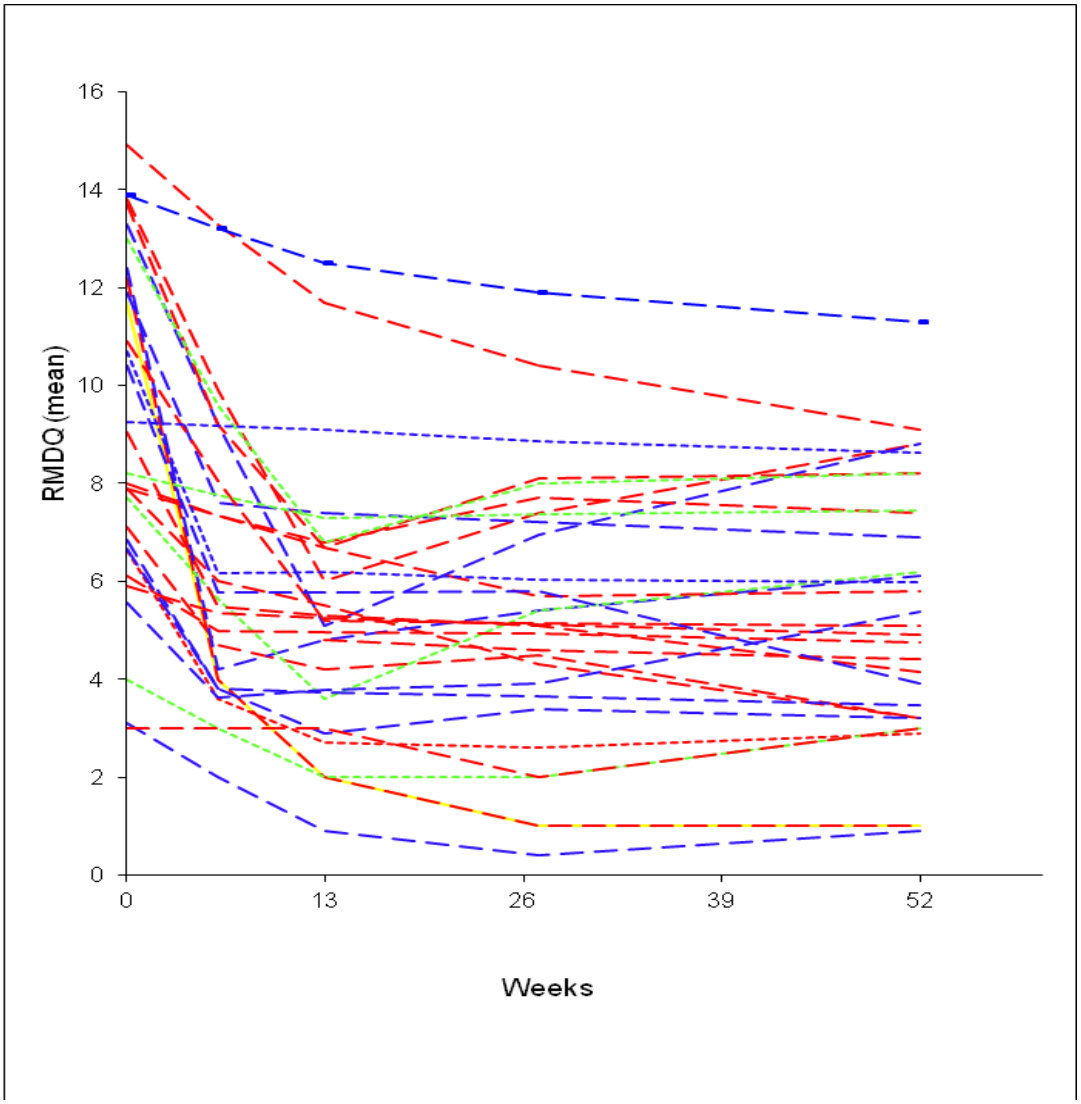


Figure 4.2b Change in RMDQ scores up to 52 week follow up in each treatment arm of included trials.
Each line represents a response line within each trial arm.
Red: Index treatment arm, Blue: Active treatment arm, Green: Usual care/waiting list/placebo arms.
____: Pharmacological treatment, - - - -: non-pharmacological treatment,
…….: Mixed/other.

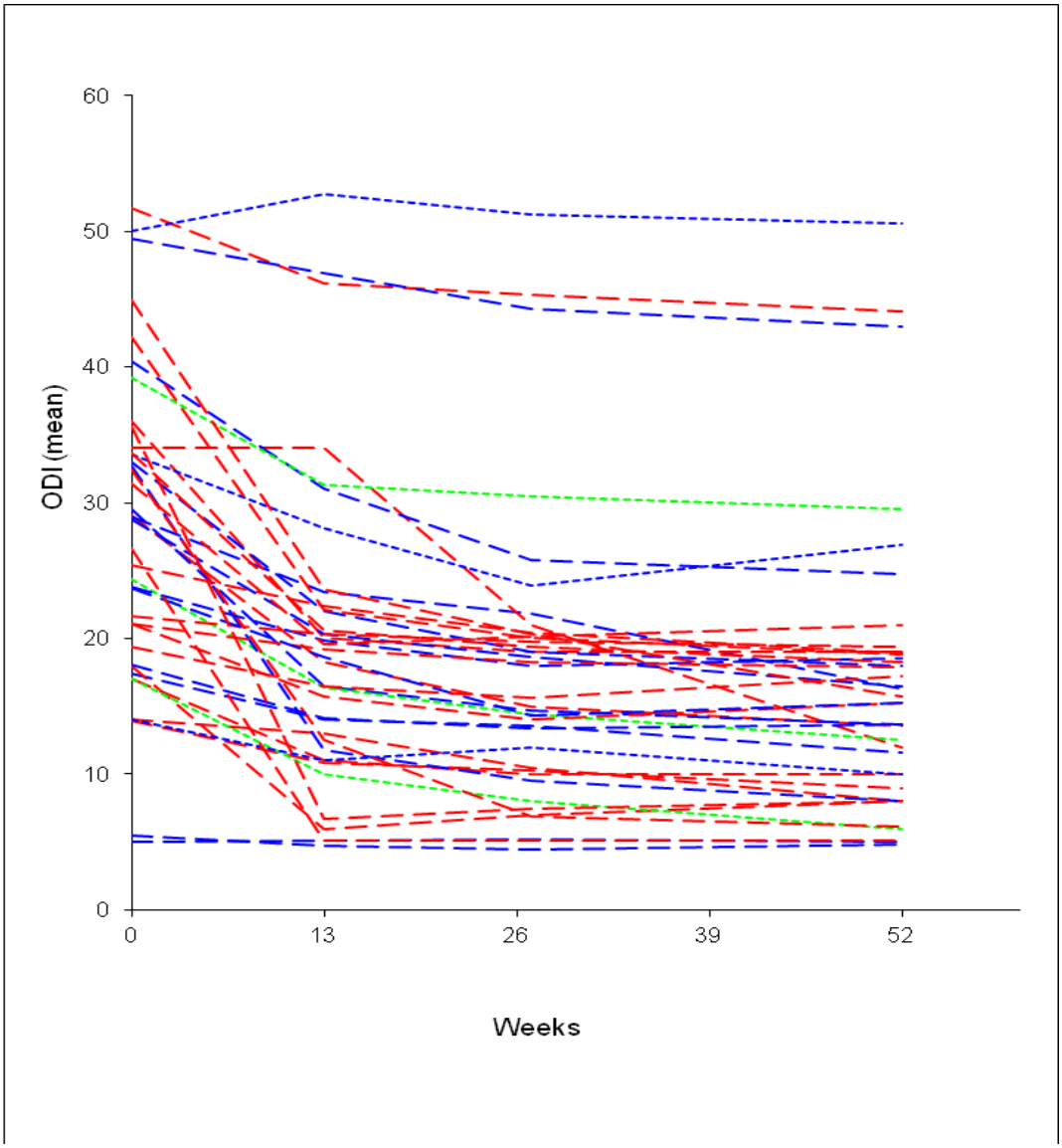


Figure 4.2c Change in ODI scores up to 52 week follow up in each treatment arm of included trials.
Each line represents a response line within each trial arm.
Red: Index treatment arm, Blue: Active treatment arm, Green: Usual care/waiting list/placebo arms.
____: Pharmacological treatment, - - - -: non-pharmacological treatment,
…….: Mixed/other.

4.4.4 Variation in the sizes of symptom progression

Figure 4.3 shows a forest plot representing SMCs and 95% confidence intervals for pain intensity at 27 week follow up for the treatment arms of the included trials, as an example of responses at other points. It shows a wide heterogeneity in the sizes of responses between trial arms (I^2-statistic = 90%). However, the majority of trial arms showed large improvement in symptoms.

4.4.5 Summary of symptom progression

Pooled SMCs were calculated based on one randomly selected arm from each trial to explore overall trend in changes in pain and functional disability over time (Table 4.1a, b). The common pattern of responses demonstrated earlier was confirmed by large initial SMCs at 13 weeks with minimum further change at 52 weeks. For pain, pooled SMC was 1.07 (95% CI 0.87, 1.27) at 13 weeks, 1.03 (95% CI 0.82, 1.25) at 27 weeks and 0.88 (95% CI 0.60, 1.11) at 52 weeks. Overall responses were large, as SMCs over 0.8 are considered as large, 0.5 – 0.8 moderate and less than 0.5 small (Cohen 1997). Results when all arms were analysed were similar to those when only one randomly selected arm per trial (Table 4.1.b). Although outcome data from trials shorter than 10 weeks were not examined in the main analyses throughout the thesis, examining these data, as presented in Tables 4.1a, b, showed that symptom progression pattern is such that the large improvement in symptom starts much sooner than 13 weeks the earliest follow point selected here.

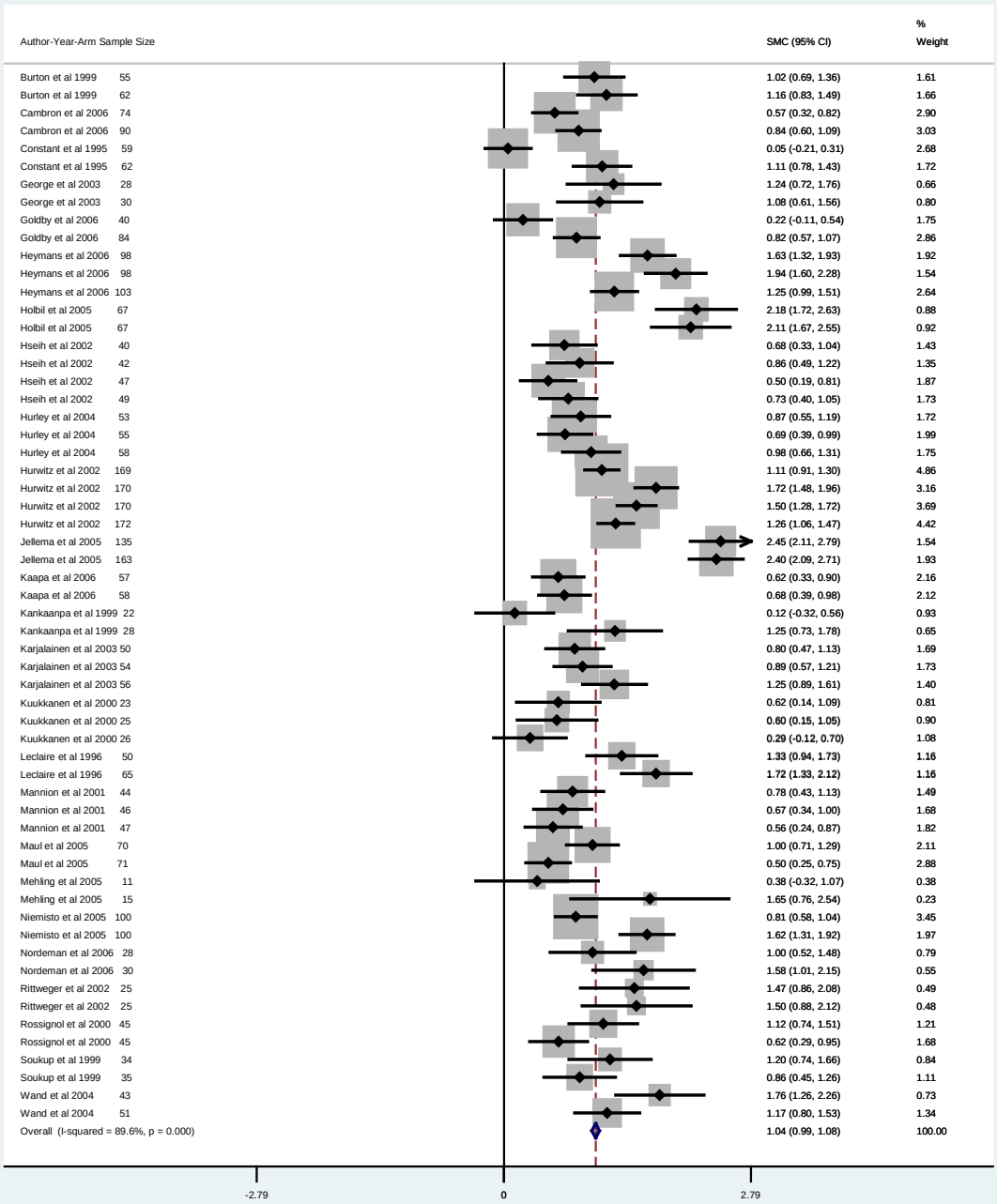


Figure 4.3 Standardised mean change (SMCs) for pain intensity in trial arms at 27 week follow up. Number of arms 50, I^2 = 90% and estimate of between-arms variance (squared tau) = 0.2421

Table 4.1a Pooled standardised mean change (SMC) for pain and disability (RMDQ and ODI) in randomly selected single arm from each trial for the follow up time points provided in the trials.

Outcome	Follow up Weeks	Trials n	Pooled SMC (95% CI)	I^2
Pain	1	13	1.44 (0.98, 1.89)	100%
	6	13	0.86 (0.65, 1.07)	82%
	13	29	1.07 (0.87, 1.27)	91%
	27	25	1.03 (0.82, 1.25)	90%
	52	23	0.88 (0.60, 1.11)	92%
	104	5	0.59 (0.45, .74)	91%
RMDQ	1	8	0.97 (0.75, 1.19)	69%
	6	19	0.97 (0.66, 1.28)	95%
	13	21	0.93 (0.67, 1.20)	94%
	27	12	0.91 (0.59, 1.24)	93%
	52	11	1.01 (0.68, 1.34)	92%
ODI	1	5	0.92 (0.59, 1.24)	58%
	6	14	0.98 (0.62, 1.33)	90%
	13	12	0.92 (0.70, 1.14)	81%
	27	10	1.08 (0.80, 1.36)	83%
	52	12	1.14 (0.88, 1.39)	84%
	104	4	1.05 (0.57, 1.54)	90%

Table 4.1b Pooled standardised mean change for pain and disability (RMDQ and ODI) for all trial arms for the follow up time points provided in the trials.

Outcome	Follow up time, weeks	Number of trials	Total number of arms	Pooled SMC (95% CI)	I^2
Pain	1	13	35	1.25 (1.13, 1.37)	100%
	6	13	33	0.83 (0.70, 0.96)	82%
	13	29	67	0.99 (9.86, 1.11)	91%
	27	25	59	1.07 (0.93, 1.20)	90%
	52	23	55	0.91 (0.76, 1.05)	89%
	104	5	13	0.77 (0.50, 1.03)	87%
RMDQ	1	8	18	0.80 (0.63, 0.98)	81%
	6	19	47	0.87 (0.77, 1.06)	95%
	13	21	45	0.86 (0.57, 1.05)	95%
	27	12	30	0.97 (0.73, 1.21)	95%
	52	11	22	0.98 (0.73, 1.23)	93%
ODI	1	5	12	0.72 (0.54, 0.89)	53%
	6	14	32	0.86 (0.65, 1.08)	89%
	13	12	29	0.73 (0.56, 0.89)	87%
	27	10	23	0.95 (0.77, 1.12)	83%
	52	12	28	0.99 (0.80, 1.19)	89%
	104	4	10	0.93 (0.66, 1.21)	87%

4.4.6 Symptom progression and types of treatments

Table 4.2 shows comparable sizes of overall responses in trial arms using index, active comparator, usual care, placebo treatments or waiting list control and also pharmacological and non-pharmacological treatments. The SMCs appear to be similar for different types of treatments. However, there is a large difference in the number of arms in groups and very small number of pharmacological and placebo treatment arms. For this reason it would not be possible at this stage to be confident of these findings which might be due to random variation. The significance of the association between types of treatment and responses to treatments will be formally tested in Chapter Six.

Table 4.2 Pooled standardised mean change (SMCs) (95% CI) for pain for trial arms stratified by type of treatment.

	13 weeks		27 weeks		52 weeks	
Types of treatments	**n***	**SMC**	**n***	**SMC**	**n***	**SMC**
Index	29	0.81 (0.75, 0.86)	25	0.95 (0.89, 1.02)	23	0.77 (0.70, 0.84)
Active comparator	25	0.81 (0.75, 0.86)	23	1.04 (0.97, 1.10)	21	0.85 (0.78, 0.92)
waiting list	3	1.33 (1.04, 1.62)	3	1.13 (0.89, 1.36)	3	0.64 (0.39, 0.89)
Placebo	2	1.56 (1.40, 1.72)	-		1	-
Usual care	8	1.15 (1.04, 1.27)	7	1.30 (1.74, 1.42)	7	1.04 (0.80, 1.18)
Pharmacological	4	0.68 (0.57, 0.79)	2	0.61(0.39, 0.83)	.	.
Non-pharmacological	53	0.91 (0.87, 0.96)	49	1.07 (1.03, 1.12)	47	0.86 (0.81, 0.91)

* Number of trial arms.

4.5 Discussion

The main aim of this review was to examine overall low back pain symptom progression in clinical trials to assess the evidence for whether they follow a similar pattern regardless of the treatment used.

Responses to treatments were examined in trials conducted over a period of 15 years in more than 18 countries among patients with back pain that varied in its duration and severity. A wide variety of primary care treatments were used ranging from tablet medications and simple advice to hands-on manual therapies, psychological treatments and extensive multidisciplinary pain management programmes.

Evidence was found that these responses follow a common pattern of early rapid and large improvement in symptoms that slows down and reaches a plateau around 13 weeks after the start of treatment. A similar pattern was found following all treatments, regardless of whether it was an index treatment, active comparator, usual care or placebo treatment, pharmacological or non-pharmacological.

4.5.1 The influence of non-specific factors

Symptom progressions in clinical trials will be influenced by the effect of the particular treatment used (the specific effect). The natural history of the symptoms, regression towards the mean, random variation, differences in trial design or analysis would also contribute to responses in clinical trials. Regression towards the mean (Morton et al 2005) is described as a statistical phenomenon where high scores or measurements outside the population mean

are likely to return and regress towards the mean over time. This could be an explanation for the large improvement in symptoms early after the start of treatment, in particular among those patients with the most severe symptoms at the start of treatment. However, this should not be seen as a mere statistical or measurement phenomenon, otherwise it would have been described as an analysis or measurement error. In fact, and particularly in the context of symptoms progression, it represents the influence of factors similar to those that influence the natural history and clinical course of symptoms. Regression towards the mean, therefore, is another term or name that refers to the same concept studied in this thesis and is not distinct from the context effect.

Providing evidence for the natural history of back pain is difficult but systematic reviews of observational studies (Hestbaek et al 2003, Pengel et al 2003) have attempted to examine its clinical course, defined as the development of a condition in the presence of treatment (von Korff 1994)) and showed it to be characterized by rapid improvement in symptoms within the first three months after inclusion that became more gradual thereafter. This echoes the pattern of responses found in this thesis.

The importance of the influence of non-specific factors is that some of them might explain the common pattern of improvement, such as being enrolled in a trial or the attention given in the trial. On the other hand, some might explain the heterogeneity between responses, such as differences in expectations regarding treatment or differences in aspects of patient-practitioner interaction.

The exact size of the influence of the non-specific factors on symptom progression is not known. Whether it is larger than the specific effect of the treatments would be interesting to explore. The size of the overall responses in clinical trials was found to be large and was the same regardless of the type of treatments used, suggesting a possibly large influence of the natural history of back pain and other non-specific factors on symptom progression.

4.5.2 Mean vs individual responses to treatment

Data on responses to treatments in clinical trials represent aggregated responses of their individual participants. Aggregated data tend, by definition, to homogenize variations in responses at individual level. Back pain trials would likely include heterogeneous groups of participants. This could lead to a potentially wide variation in the progression of their symptoms in the trials that would not be well represented by the aggregated data presented in published data reports which were used in this review. This wide variation in individual symptom progression is one basis for the proposals of stratified care approaches that have been developed more recently to target more homogenous sub-groups of patients with similar characteristics and who represent different prognostic subgroups (Hay et al 2008), or who may respond better to specific types of treatment (Fritz & George 2000).

4.5.3 Large change in outcome scores over time

An important finding from this review is the large symptom progression common in all trial arms, active as well as placebo, usual care or waiting list arms. It is interesting that evidence exists for a large overall improvement in back pain symptoms in all arms of clinical trials, while more and more trials are unable to

show clear evidence for the efficacy (specific effects) of the active treatments. Randomized controlled trials are designed to attribute the benefit of a treatment solely to the difference between its effect and that of the comparator treatment. Such explanatory trials are important and necessary to study the efficacy of treatments. However, it is also important to explore the role of other factors, in addition to the specific ingredient of the treatment used. Addressing these non-specific factors, which may include factors related to the patients, practitioners, the setting or the way treatments are designed and delivered, may help us better understand the overall symptom progression in clinical trials.

4.6 Limitations

This review provides empirical evidence for the hitherto anecdotal observation increasingly being made that randomised clinical trials on primary care treatments for NSLBP show modest or non-significant differences between various types of treatments, but large within group changes over time. A large number of trials was included, on a wide variety of treatments, pharmacological and non-pharmacological, manual and psychological. A number of limitations, nevertheless, merit addressing.

4.6.1 The choice of data source

Restricting the source of data to the CENTRAL database might have limited the overall number of trials included. However, this database satisfied the aims of this review, namely providing a large pool of clinical trials on a wide range of primary care treatments for NSLBP to explore the evidence for a common pattern for the

overall responses to treatments. It is arguably a strength of the CENTRAL database that it does not include only trials registered in MEDLINE and EMBASE, but also trials identified from hand searches and trials added by speciality groups, including the Back Review Groups. It is, therefore, not expected that a wider search would have yielded a much larger pool of trials. However, it would certainly be interesting to update the review with the expectation of capturing more trials which would add power to its outcomes.

Only trials published in English were included. It may be argued that this language restriction could lead to selection bias as the context effect may be influenced by local cultural factors or differences in the delivery or quality of health care. However, there are a number of trials included which were published in English even though they were conducted among non-English-speaking populations. Therefore, the review does include trials from a variety of countries and cultures. Furthermore, the evidence on the benefit or otherwise of including trials published in languages other than English in systematic reviews is conflicting (Egger et al 1997, Gregoire et al 1995, Juni et al 2002, Moher et al 2000). There was no other evidence for selection bias as the included trials covered a wide range of treatments in a wide range of primary care settings with no evidence of systematic lack of a particular group or type of trials.

4.6.2 Choice of outcome measures

Although the initial intentions were to include as many outcome measures as possible, outcome measures were analysed that were most commonly used in order to include a sufficiently large number of trials. A small number of trials used other measures such as patient's global perceived effect or measures to assess

depression or return to work, but these outcome measures were not widely used and/or did not have standardized definitions or scales that would allow meaningful comparison. Hence, although these measures would add to describing the totality of patient experience and provide an important representation of patient symptom progression, they were not included in the analysis. Recommendations have been made towards a standard group of core outcome measures for use in low back pain trials (Dworkin et al 2005), which would facilitate comparison and combination of results.

4.6.3 Other methodological limitations

Although the random effects model was used in the meta-analysis, pooling response sizes with a high heterogeneity is still problematic. However, it is important to emphasize that the purpose of pooling for this review was merely to further assess and present the pattern of overall responses rather than to calculate estimates of particular treatment effect size. Furthermore, trials on a wide variety of disparate types of treatments were included and therefore even if heterogeneity were low, the result of such pooling would obviously not have been clinically or practically meaningful. Any conclusions drawn from pooling in this review, therefore, should be made within the context of the particular purpose for using it here.

4.7 Summary

Overall responses to treatments in NSLBP clinical trials were shown to be large and follow a common pattern of rapid early improvement followed by a plateau following all types of treatments.

Given such a similar pattern of responses, it is not surprising that any specific effect of treatment would be difficult to detect. It is important to explore factors that are associated with symptom improvement in clinical trials. That might be associated with any variation in the size of responses. Duration of symptoms (acute, subacute or chronic), severity of symptoms, patients' age, gender, preference for treatments, expectations and patient-practitioner interaction are examples of factors that might influence the progression of symptoms in individual patients. Identifying the association between responses to treatments and such factors would be the first step towards identifying subgroups of patients and targeting of treatments for these groups of patients with common factors or utilizing and harnessing the influence of these factors to improve patient outcome in clinical trials. In Chapter Six, results of meta-regression analyses to examine associations between such trials' factors and responses to treatments will be outlined.

Exploring responses to treatments for pain conditions other than low back pain, and indeed for other medical conditions, was beyond the scope of this review. It would be interesting to explore whether these findings are unique to NSLBP and therefore reflect in part the poorly understood nature of this condition; or if they

could be reproduced in other medical conditions, raising wider issues about how treatment outcomes and symptom progression in general are assessed.

Chapter Five

Quality of randomised clinical trials of primary care treatments for non-specific low back pain

5.1 Introduction

Assessing the quality of clinical trials is accepted as an important part of any systematic review (Chalmers 1989, Oxman et al 1991, Cook et al 1995). Reviews rely, for a large part, on the quality of their primary trials to support the robustness and validity of their conclusions (Kassirer et al 1994). Moher et al (1999) pointed out that the validity of the outcomes of systematic reviews and meta-analyses draws on the ability to demonstrate that the included trials followed valid methods and were conducted according to a clear and correct protocol. The quality of trials included in the systematic review in the previous chapter was therefore assessed to ensure that the conclusions regarding a common pattern of improvement in back pain trials were based on sound evidence.

Certain aspects of trial methods and conduct might influence responses to treatments. Empirical evidence indicates that poor conduct of some methodological aspects influence estimates of treatment effect such as inadequate methods of random-sequence generation, allocation concealment, or blinding (Moher et al 1998, Schulz et al 1995, Kjaergard et al 2001). Assessing the quality of the RCTs included in the review would therefore provide an opportunity for examining the association with responses to treatments in these trials. Such analysis was undertaken and the results are presented in Chapter Six.

A large number of methods, checklists and scales have been used to assess trials' quality (Moher et al 1995, Verhagen et al 2001). As the interest in this thesis was in responses to treatments in clinical trials, rather than treatment effect or the generalisability of trials' outcomes, the focus was on assessing the internal validity

of the trials. At the time of conducting this study, a quality checklist was recommended and used by the Cochrane Back Review Group (van Tulder et al 2003) and was used. It is a composite tool that assesses 11 individual quality criteria related to the internal validity. The overall quality of the trial is evaluated by calculating the summary score based on the scores (positive, negative, or unclear) for the individual criteria.

Information on some quality criteria can be unavailable in the published reports. Considering such criteria as negative or unsatisfactory might lead to inaccurate description of the quality of the trial for it mixes defective conduct with poor reporting (Hill et al 2002). Both poor conduct and poor reporting represent serious defects. However, poor conduct could arguably be more significant than poor reporting and mixing the two might not, therefore, be ideal. The aim in this chapter was to assess the quality of randomised controlled trials with a particular focus on the methods of assessing the overall quality and individual quality aspects of the trial and the issue of lack of information in the published reports.

5.2 Objectives

The main objective was to assess the quality of RCTs on primary care treatments for non-specific low back pain that were included in the systematic review described in Chapter Four. In using a quality assessment check list of individual quality criteria, the summary versus individual quality criteria scores were also assessed and compared.

The secondary objectives were to explore the association between trials' quality and other trials' characteristics including their setting and type of treatment (pharmacological and non-pharmacological). The extent of missing information in published reports of trials and the impact this might have on the overall outcome of quality assessment was also assessed. Finally, whether trial quality or availability of information in the published reports improved since the CONSORT statement was published in 1996, was also explored.

5.3 Methods

5.3.1 Data source

The Cochrane Register of Controlled Trials (CENTRAL), the first quarter issue of 2007 accessed in April, was the source of the trials included in the review. Details of inclusion and exclusion criteria were described in detail in Chapter Four.

5.3.2 Assessment of methodological quality

The quality assessment criteria checklist of the Cochrane Back Review Group (CBRG) (van Tulder et al 2003) (Appendix 5) was modified and used to assess the internal validity of the selected trials. The modifications, which are detailed here, were mainly made to take into account the main aims and objectives of this review. To clarify the modifications, the numbers of the modified criteria were marked with the letter 'm'. The final checklist that was used is presented in table 5.1.

Table 5.1 The quality criteria checklist used for this review modified from the Cochrane Back Review Group checklist

1. Was the method of randomization adequate?
2. Was the treatment allocation concealed?
3. Was the patient blinded to the intervention?
4. Was the care provider blinded to the intervention?
5. Were co-interventions prevented/avoided?
6. Were co-interventions standardised?
7. Was compliance acceptable?
8. Was the drop-out rate acceptable?
9. Was the timing of the outcome assessment in all groups comparable?
10. Was the analysis based on intention-to-treat analysis?

Item (3) '*Were the groups similar at baseline regarding the most important prognostic factors?*' was dropped as the focus in the review was to assess within-group symptom progression rather than between group differences. Item (6) '*Was the outcome assessor blinded to the intervention?*' was also dropped as all the outcome measures selected for the review were subjective patient completed measures.

The item on co-interventions (7), was split to deal specifically with the issue of whether particular instructions were given to participants on 'standardisation of co-interventions' (7m) in addition to avoiding using such interventions (6m). Consequently, when item (6m) was scored yes, i.e. instructions were given to avoid using co-interventions, item (7m) was automatically scored 'not applicable'.

Subgroup analyses were conducted to assess differences in overall trial quality and individual criteria scores according to subgroups of trials based on setting (general practice, physiotherapy units, occupational healthcare units, general population, mixed and unclear setting), type of treatment (pharmacological vs non-pharmacological) and publication date (before or after 1996). The purpose of

classifying trials according to the type of treatment was to explore the often described difficulty in satisfying some of the quality criteria in trials on non-pharmacological treatments, such as patient and practitioner blinding (Black 1996, Boutron et al 2003, 2004).

The author and a second reviewer (DvdW) independently assessed the quality of trials. The criteria were scored "yes" (item fulfilled), "no" (item not fulfilled) or "don't know" (insufficient information). It was the reports of the trials that were assessed and criteria scored positive or negative only when the published information was clear, otherwise a ‘don’t know’ score was given. The assessment was unblinded as author, institution, journal names, or trial results were not concealed. Meetings were organised to discuss disagreements and obtain consensus on the final score.

5.3.3 Data analysis

Quality assessment data were analysed for the 10 individual criteria and also summed up into a total summary score, which was calculated based on the number of positively scored items, with equal weights of one. An arbitrary 50% cut-off point is commonly used to indicate high quality (van Tulder et al 2003). Therefore, for this review high quality was defined as having five or more positive scores. Data on the ‘don’t know’ scores, representing missing information, were analysed separately and the consequences for quality assessment of merging them with either negative or positive responses were also explored. The significance of differences between subgroups was analysed using chi-squared tests

5.4 Results

5.4.1 Trials' description

As reported in Chapter Four, the search yielded a total number of 772 citations. One hundred and twenty six satisfied the inclusion criteria. Eight published papers were each a second report of the same trial. Multiple publications from a same trial were considered as one source and used to obtain information to assess the paper quality as well as outcome data for the purpose of the systematic review. Data, therefore, were available for 118 trials. Further detailed description of the included trials is provided in Appendix 4.2.

5.4.2 Quality assessment

The two authors who independently assessed trials' quality agreed on similar scores for all quality criteria for 15 trials (13%). Initial disagreements for the remaining 103 trials (87%) ranged from disagreement on scoring one criterion (for 17 trials, 14%) to seven criteria (for one trial only). Agreement was reached on 883 items out of a total of 1180, (75%). Disagreements were mainly on the criteria on co-interventions and intention-to-treat analysis. All disagreements were resolved through consensus.

5.4.3 Summary score analysis: overall trials' quality

Just under half of the included RCTs were of high quality using the definition of at least five positive scores (56, 48%). There was a wide variation in the quality of the included trials, from trials with only one positive score, five negative scores and four 'don't know' to trials with information available on all criteria, scoring eight positive and two negative scores. The median number of positive scores given for

each trial was four, range 1, 8; of negative scores 3, range 0, 7; and of ‘don’t know’ scores 2, range 0, 8.

5.4.4 Individual criteria assessment

A summary of the scores for the individual quality criteria is presented in Table 5.2 and Figure 5.1. Among the criteria for which information was available, the criteria that were found to be most commonly satisfied i.e. scored ‘yes’, were ‘was timing of outcome measurement in all groups comparable’ (94%), ‘was dropout rate acceptable’ (65%) ‘was the analysis based on intention to treat analysis’ (62%), ‘was the method of randomisation adequate’ (61%) and ‘was the treatment allocation concealed’ (55%). The criteria that were least commonly satisfied in the trials, i.e. scored ‘no’, were ‘was care provider blinded’ (78%) and ‘were patients blinded’ (59%).

Although information was available for most quality criteria in the majority of trials included in this review, information was not available for at least one quality criterion in 98 trials (83%). ‘Don’t know’ scores ranged from only one trial for the criterion on ‘timing of outcome measurement’ to 55 trials (47%) for the criterion on ‘compliance’.

Table 5.2 Number (percentage) of trials scored positive, negative or 'don't know' for each individual quality criterion.

Quality criteria	Yes N (%)	No N (%)	Don't know N (%)
Randomisation adequate	72 (61)	10 (9)	36 (31)
Allocation to treatment concealed	65 (55)	1 (0.1)	52 (44)
Patient blinded	31(26)	69 (59)	18 (15)
Care-provider blinded	15 (13)	92 (78)	11 (9)
Co-interventions prevented	26 (22)	52 (44)	40 (34)
Co-interventions standardised [a]	10 (11)	38 (41)	44 (48)
Compliance acceptable	53 (45)	10 (9)	55 (47)
Drop out acceptable	77 (65)	34 (29)	7 (6)
Timing of measurement comparable [b]	111 (94)	6 (5)	1 (0.1)
Intention to treat analysis	73 (62)	21 (18)	24 (20)

[a] This was not applicable in 26 trials; [b] Timing of measurement of outcomes was comparable across all subgroups in the trial.

Adding the 'don't know' scores to either the negative or positive scores would have the largest impact for criteria for which information was not available in a large number of trials, such as prevention and standardisation of co-interventions, compliance and concealment of allocation to treatments, as shown in Table 5.3. When the 'don't know' scores were added to the negative scores, the majority of trials scored negative for these criteria, whilst when the scores were added to the positive scores the majority of trials scored positive for these criteria.

Table 5.3 Number (percentage) of trials scored positive and negative for each quality criterion when 'don't know' scores were added to the negative and next to the positive scores

	'Don't know' added to 'No' scores		'Don't know' added to 'Yes' scores	
Quality criteria	**Yes n (%)**	**No n (%)**	**Yes n (%)**	**No n (%)**
Randomisation adequate	72 (61)	46 (39)	108 (91)	10 (9)
Allocation to treatment concealed	65 (55)	53 (45)	117 (99)	1 (0.1)
Patient blinded	31(26)	87 (74)	49 (41)	69 (59)
Care-provider blinded	15 (13)	103 (87)	26 (22)	92 (78)
Co-interventions prevented	26 (22)	92 (78)	66 (56)	52 (44)
Co-interventions standardised [a]	10 (11)	82 (89)	54 (59)	38 (41)
Compliance acceptable	53 (45)	65 (55)	108 (91)	10 (9)
Drop out acceptable	77 (65)	41 (35)	84 (71)	34 (29)
Timing of measurement comparable [b]	111 (94)	7 (6)	112 (95)	6 (5)
Intention to treat analysis	73 (62)	45 (38)	97 (82)	21 (18)

[a] This was not applicable in 26 trials; [b] Timing of measurement of outcomes was comparable across all subgroups in the trial.

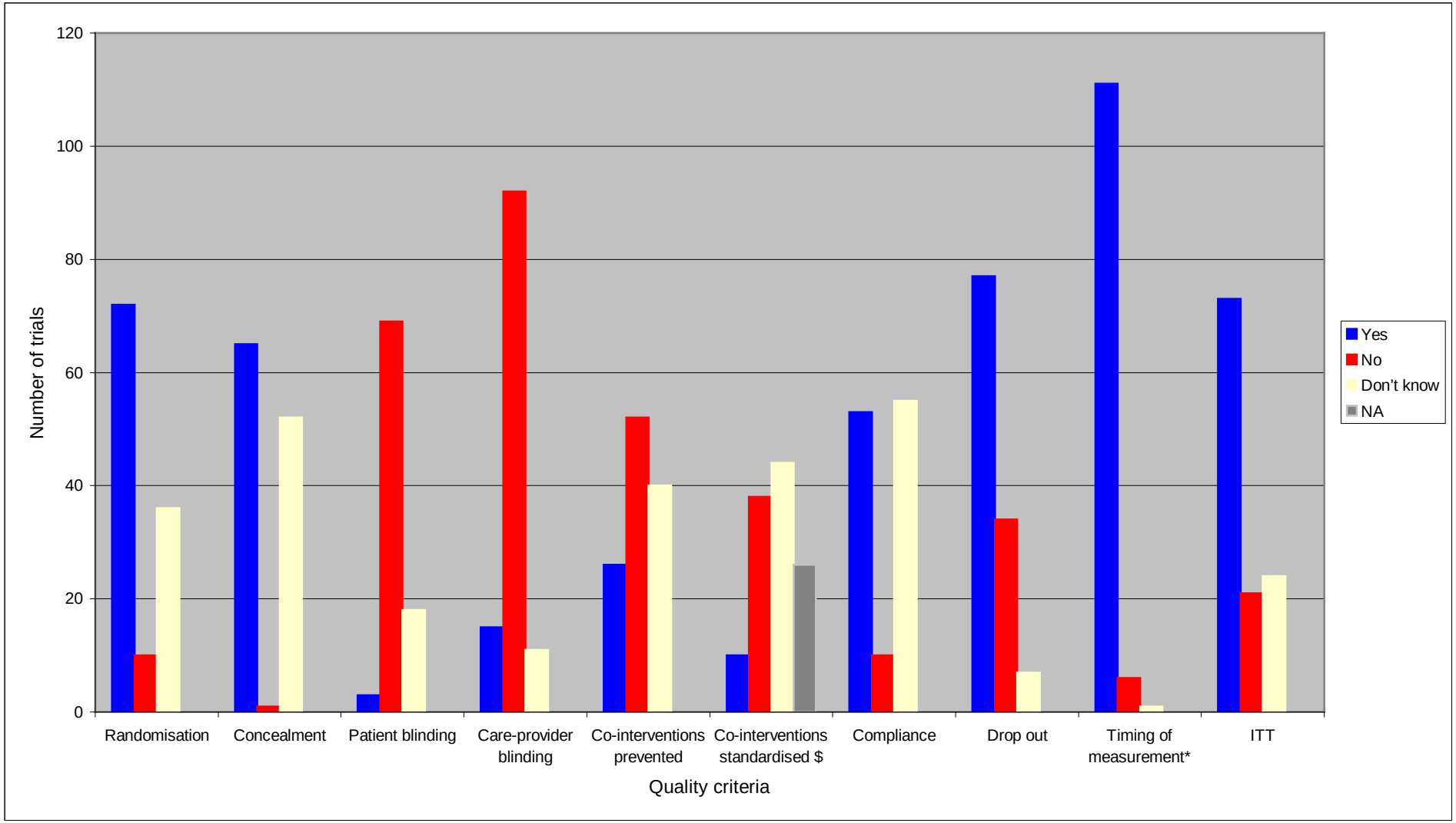


Figure 5.1 Distribution of total scores for quality assessment criteria for included trials

* Was timing of measurement of outcomes comparable between groups. ITT: Analysis was on basis of intention to treat.

5.4.5 Trials' quality and other trials' characteristics

The association between quality and trials' characteristics such as setting (primary care, occupational healthcare, physiotherapy, general population, mixed and unclear setting), type of intervention (pharmacological or non-pharmacological) and publication dates was examined and the results are outlined below:

5.4.5.a Trials' quality and setting

Twenty-nine trials were conducted in primary care / family medicine setting, 20 in occupational healthcare departments, 10 in physiotherapy / physical therapy units and 14 were conducted among the general population. 31 trials were conducted in 'mixed setting' including a primary care setting and in 14 trials the setting was not clear.

Tables 5.4a & b, show the distribution of scores for individual quality criteria according to trials' setting. The lack of information was widespread with statistically significant difference between types of setting (Table 5.5), with information most widely missing in trials in physiotherapy setting and least in general practice and mixed setting trials. The number of trials that lacked information on at least one quality criterion ranged from seven trials in physiotherapy setting (70%) to 14 trials for which the setting was not clear (100%).

Table 5.4a Number (percentage) of trials scored positive, negative and 'don't know' for each quality criterion according to trials' setting.

	General practice setting N = 29			Occupational Health Setting N = 20			Physiotherapy Setting N = 10		
Quality criteria	**Yes N (%)**	**No N (%)**	**Don't know N (%)**	**Yes N (%)**	**No N (%)**	**Don't know N (%)**	**Yes N (%)**	**No N (%)**	**Don't know N (%)**
Randomisation adequate	20 (69)	1 (5)	8 (26)	14 (70)	3 (15)	3 (15)	4 (40)	3 (30)	3 (30)
Allocation to treatment concealed	17 (59)	0	12 (41)	11 (55)	0	9 (45)	4 (40)	0	6 (60)
Patient blinded	6 (21)	17 (59)	5 (17)	1 (5)	18 (90)	1 (5)	2 (20)	6 (60)	2 (20)
Care-provider blinded	3 (10)	22 (76)	3 (14)	0	19 (95)	1 (5)	0	9 (90)	1 (10)
Co-interventions prevented	7 (24)	15 (52)	7 (24)	3 (15)	8 (40)	9 (45)	0	3 (30)	7 (70)
Co-interventions standardised	2 (9)	13 (59)	7 (32)	0	6 (35)	11 (65)	1 (10)	2 (20)	7 (70)
Compliance acceptable	19 (66)	1 (5)	9 (29)	7 (35)	4 (20)	9 (45)	2 (20)	0	8 (80)
Drop out acceptable	18 (62)	8 (28)	3 (14)	13 (65)	6 (30)	1 (5)	6 (60)	3 (30)	1 (10)
Timing of measurement comparable*	28 (97)	1 (5)	0	18 (90)	2 (10)	0	10 (100)	0	0
Intention to treat analysis	16 (55)	5 (17)	8 (26)	11 (55)	3 (15)	6 (30)	7 (70)	1 (10)	2 (20)
Median score per trial	4	3	2	4	3	2	2	2	2

*: Timing of measurement of outcomes was comparable across all subgroups in the trial.

Table 5.4b Number (percentage) of trials scored positive, negative and 'don't know' for each quality criterion according to trials' setting.

	General population setting N = 14			Mixed setting N = 31			Setting unclear N = 14		
Quality criteria	**Yes N (%)**	**No N (%)**	**Don't know N (%)**	**Yes N (%)**	**No N (%)**	**Don't know N (%)**	**Yes N (%)**	**No N (%)**	**Don't know N (%)**
Randomisation adequate	8 (57)	1 (7)	5 (36)	22 (71)	2 (7)	7 (22))	4 (29)	0	10 (71)
Allocation to treatment concealed	8 (57)	0	6 (43)	18 (58)	1 (3)	12 (39)	7 (50)	0	7 (50)
Patient blinded	4 (29)	5 (36)	5 (36)	6 (19)	21 (68)	4 (13)	10 (71)	2 (14)	2 (14)
Care-provider blinded	1 (7)	12 (86)	1 (7)	4 (13)	26 (84)	1 (3)	6 (43)	4 (29)	4 (29)
Co-interventions prevented	4 (29)	8 (57)	2 (14)	5 (16)	14 (45)	12 (39)	7 (50)	4 (29)	3 (21)
Co-interventions standardised	2 (20)	5 (50)	3 (30)	2 (8)	11 (42)	13 (50)	3 (43)	1 (14)	3 (43)
Compliance acceptable	7 (50)	0	7 (50)	14 (45)	5 (16)	12 (39)	4 (29)	0	10 (71)
Drop out acceptable	13 (93)	0	1 (7)	19 (61)	12 (39)	0	8 (57)	5 (36)	1 (7)
Timing of measurement comparable*	14 (100)	0	0	28 (90)	2 (7)	1 (3)	13 (93)	1 (7)	0
Intention to treat analysis	9 (64)	2 (14)	3 (21)	23 (74)	4 (13)	4 (13)	7 (50)	6 (43)	1 (7)
Median score per trial	5	2	1	4	3	2	4	1	3

*: Timing of measurement of outcomes was comparable across all subgroups in the trial

Table 5.5 Distribution of the 'don't know' scores according to setting and type of treatment

	Number of trials	Total number of criteria scored	Criteria scored unclear, n (%)
Setting			
General practice	29	290	62 (21)
Occupational healthcare	20	200	50 (25)
Physiotherapy	10	100	37 (37)
General population	14	140	32 (23)
Mixed	31	310	66 (21)
Unclear	14	140	41 (29)
			P = 0.018
Type of treatment			
Pharmacological	20	200	44 (22)
Non-pharmacological	91	910	233 (26)
			P = 0.286

Using a 50% cut-off point for the summary score analysis, the number of high quality trials was 16 (55%) in general practice setting, eight (40%) in occupational healthcare setting, three (30%) in physiotherapy setting, eight (57%) in general population setting, 15 (48%) in mixed setting and seven (50%) in trials in which the setting was not clear. The differences in the percentages of high quality trials in these settings were not significant (Pearson chi-square <df=5 > = 2.896, p=0.716).

5.4.5.b Trials' quality and type of treatment (Table 5.6)

Ninety-one trials evaluated the effectiveness of non-pharmacological treatments, 20 trials pharmacological treatments and five trials mixed treatments and it was not possible to classify two trials according to the type of treatment.

Table 5.6 Number (percentage) of trials scored positive, negative and 'don't know' for each quality criterion according to type of intervention.

	Pharmacological interventions N = 20			Non-pharmacological interventions N = 91		
Quality criteria	**Yes N (%)**	**No N (%)**	**Don't know N (%)**	**Yes N (%)**	**No N (%)**	**Don't know N (%)**
Randomisation adequate	13 (65)	0	7 (35)	54 (59)	10 (11)	27 (30)
Allocation to treatment concealed	12 (60)	0	8 (40)	50 (55)	1 (1)	40 (44)
Patient blinded	16 (80)	1 (5)	3 (15)	11 (12)	63 (69)	17 (19)
Care-provider blinded	13 (65)	0	7 (35)	0	86 (95)	5 (6)
Co-interventions prevented	11 (55)	7 (35)	2 (10)	15 (17)	39 (43)	37 (40)
Co-interventions standardised	4 (44)	3 (33)	2 (22)	4 (5)	31 (41)	41 (50)
Compliance acceptable	7 (35)	1 (5)	12 (60)	43 (47)	9 (10)	39 (40)
Drop out acceptable	13 (65)	6 (30)	1 (5)	61 (67)	24 (26)	6 (7)
Timing of measurement comparable*	20 (100)	0	0	84 (92)	6 (7)	1 (1)
Intention to treat analysis	14 (70)	4 (20)	2 (10)	55 (60)	16 (18)	20 (22)
Median score per trial	7	1	2	4	3	3

*Timing of measurement of outcomes was comparable across all subgroups in the trial

There was no difference in the number of criteria on which information was missing (scored 'don't know') between trials on non-pharmacological treatments (233 out of a total of 910, 26%) and trials on pharmacological treatments (44 out of a total of 200, 22%) (Pearson chi-square < df = 1 > = 0.696, p= 0.286,) (Tables 5.5 and 5.6). One exception was information on care-provider blinding which was lacking more often among trials on pharmacological treatments compared with non-pharmacological treatments trials (7 out of 20, 35% vs 5 out of 91, 6% respectively).

There was a larger number of high quality trials on pharmacological treatments compared with trials on non-pharmacological treatments (16/20, 80% vs 34/91, 37% respectively) (Pearson chi-square <df = 1> = 12.041, p=0.001). Among trials that provided information on care provider blinding, all trials on pharmacological treatments were given positive scores while none of the non-pharmacological

trials was given positive scores. For patient blinding, 80% of pharmacological trials were given positive scores compared with 12% of non-pharmacological trials. There was no significant difference between trials on the two types of treatments regarding positive scores on the remaining criteria.

5.4.5.c Trials' quality and year of publication (Figure 5.2)

Trials included in this review were published during a period of 14 years (1993-2007) with a median number of seven trials published per year (range 1, 20). The majority of included trials were published in the years 2002-2007 (80/118, 68%). 15 trials were published during 1993-1995, 102 during 1997-2007 and one trial in 1996.

Out of 15 trials published before the publication of the CONSORT statement in1996, 14 trials were scored 'don't know' for at least one quality criterion (93%) (median of 4 'don't know' scores per trial, range 0,6), compared with 83 out of the 102 trials published after 1996 (81%) (median of 2 'don't know' scores per trial, range 0,8). This would suggest a positive trend regarding adequate reporting of information, although the difference was not statistically significant (Fisher exact test, $p = 0.226$). The single trial that was published in 1996 was scored 'don't know' for two criteria.

The number of positive scores did not differ among trials published before and after 1996 (median of five positive scores per trial published before 1996, range

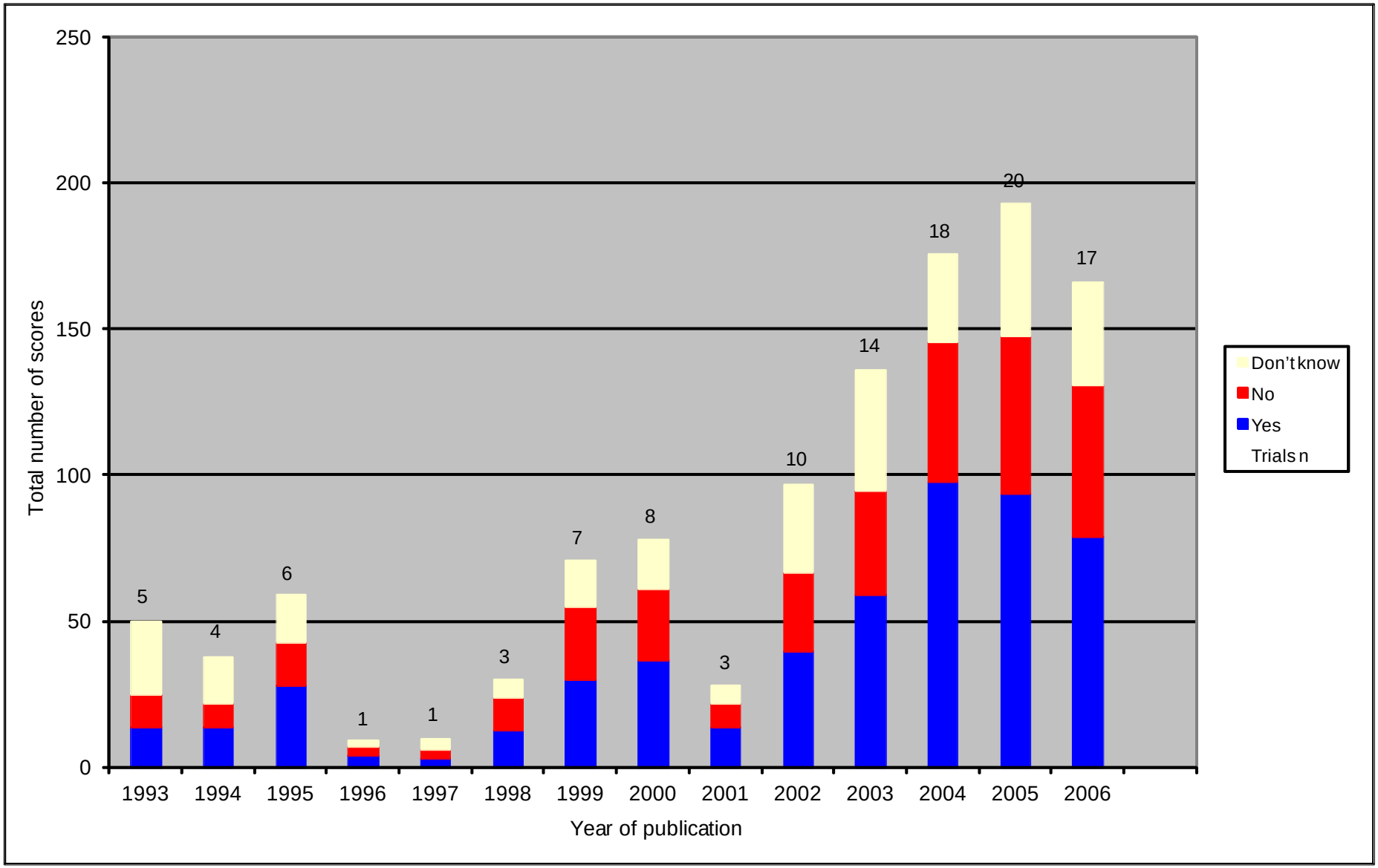


Figure 5.2 Quality scores for all included trials for each publication year

Number above bars refers to the number of trials. Year 2007: not shown as only included trials published before April 2007 (incomplete year), refer to main text on trials' inclusion.

1,8; compared with a median of five positive scores per trial published after 1996, range 1,8). The single trial published in 1996 was scored positive for four criteria.

Although it seems that more high quality trials were published after 1996 compared with before 1996 (52 out of 102, 51% vs 4 out of 15, 27% respectively), the difference between the two groups was not statistically significant (Pearson chi-square <df=1> =3.098 p=0.078). The single trial published in 1996 was considered to be of a low quality.

5.5 Discussion

Further to assessing the quality of trials as part of a systematic review, the aim was to particularly focus on the interpretation and analysis of quality data using the summary score analysis, describing individual quality criteria and inadequate reporting.

Just under a half of the trials included in this study were considered to be of high quality with the majority falling below the expected standards. This echoes the findings of a study published in 2005 (Koes et al 2005) which specifically assessed the quality of back pain RCTs using the same quality assessment tool. In that review, Koes et al assessed the quality of 269 trials included in 15 Cochrane systematic reviews of conservative treatments for low back pain over the period 1961-2004. They concluded that many of the RCTs did not meet the quality standards with a wide variation in their quality. Furthermore, and which is perhaps even more significant, they did not find evidence for improvement in

methodological quality of trials over the time from 1960 to 2004. The potential impact of the findings related to the quality of the RCTs included in the systematic review presented in Chapter Four is important to consider. This impact, or the association between quality and symptom progressions, will be formally tested in Chapter Six.

For the broader implication of the findings in this thesis, it is important to note that trials' inclusion criteria for Koes et al study (2005) are different from those of this review (Chapter Four) where RCTs were included to meet a specific objective which was to examine responses to treatments. Therefore, many trials were excluded from this review as they were concerned with either ineligible populations, treatments, outcomes or setting and their quality was not assessed. Caution therefore needs to be exercised regarding generalising the findings regarding trials quality to all back pain trials.

The methods used and quality of subgroups of trials according to setting, type of treatment and publication date are discussed below.

5.5.1 The summary score analysis

One potentially major flaw with the summary score analysis is that the cut-off point is arbitrary with equal weights given to the individual criteria. The attraction of this analysis is obvious in that it offers an easy and communicable way to describe trials as poor or high quality that could also be used to examine the association between quality and trials outcomes or other characteristics.

The summary score analysis does not have the ability to discriminate between individual criteria according to their importance or relevance. Criteria such as adequacy of random sequence generation, selective drop-out, and concealment of allocation to treatment could probably be considered as more relevant and indicative of trial quality and unbiased outcome than, for example, comparability of timing of follow up across trial groups. The CBRG checklist did not assign weights to the criteria while some quality assessment lists such as the Maastricht List did (Verhagen et al 1998), even though the weights were arbitrary and not based on universally agreed consensus.

After finishing writing the findings from this study the Cochrane Collaboration announced that they adopted a new quality assessment tool across all review groups which is the 'risk of bias' tool (Furlan et al 2009). This came as part of the Collaboration revising their whole reviewing process. The new tool addresses six specific domains: 'sequence generation', 'allocation concealment', 'blinding (patients and practitioners)', 'incomplete outcome data',' selective outcome reporting' and 'other issues'. Each of these domains is scored 'yes', 'no' or 'unclear', similar to the tool used in this thesis. The domains of sequence generation, allocation concealment and selective outcome reporting, are to be addressed for the whole trial. Other domains could be considered separately for separate outcome measures. The emphasis in the new tool is on encouraging reviewers to make judgement when drawing conclusions about the overall risk of bias. This indicates that the new tool addresses some of the caveats associated with the previous tool by trying to put different weights on various domains. More significantly, the recommendation is to keep the assessment for each domain separate and not to combine them into an overall summary score.

5.5.2 Analysis of individual quality criteria

The criteria most frequently scored negative were related to blinding of patients and care providers and also managing co-interventions (prevention or standardisation). On the other hand, in a large number of trials randomisation was considered adequate as well as concealment of allocation to treatment and reporting of drop-out rates.

This type of data provides insight into the particular weaknesses and strengths of aspects of trials' quality, compared with merely providing an overall description of the trial as high or poor quality or the overall proportion of high quality trials. Identifying particular areas in which trials are consistently inadequate, such as blinding, can be explored further to identify causes and possible solutions. Data on individual quality criteria can also help assess the impact of particularly relevant criteria on outcomes and conclusions for particular types of treatments or types of trials.

5.5.3 Quality of reporting

The decision in this study was to assess the trials through assessing their published reports. To this extent, the assessment could arguably be described as quality assessment of reporting of trials rather than of the trials themselves. Assessment of trial quality is inherently linked with the quality of their reports (Moher et al 1995) and the incomplete reporting of important aspects of trials methods is a common problem (Mostellier et al 1980, Der Simonian et al 1982, Schulz et al 1994). The question is whether poor reporting is considered as equal to poor conduct. Studies have shown that the lack of information on

methodological aspects of trials in their published reports did not always reflect real flaws in their conduct, rather a reporting flaw (Hill et al 2002), although some studies contradicted this by implying that lack of information in the report of a trial was associated with inadequate conduct (Schulz et al 1995). One possible solution is to approach authors and seek the missing information. However, in addition to the obvious practical and logistic burden this would entail, this would not guarantee complete or accurate information. It was decided to rely on published reports to assess trials quality as the argument was that systematic reviews usually rely on published reports of RCTs in compiling their evidence and it is the published reports that are linked with the trials outcomes and conclusions. Using this approach, it is important to closely examine the missing information as a distinct part of assessing the quality of trials. Nevertheless, there are strong arguments and justifications for both approaches of either relying only on published reports or contacting authors, neither of which is perfect. The perfect solution would be for trialists to publish complete information on their trials.

In 1996 the Consolidation of Standards for Reporting Trials (CONSORT) statement was published (Begg et al 1996) and revised in 2001 and 2010 (Moher et al 2001, Consort statement website). It has been endorsed by the World Association of Medical Editors, the International Committee of Medical Journal Editors (ICMJE), and the Council of Science Editors. One aim of the statement was to provide authors and publishers with guidelines on the information that should be included in RCT reports to enable readers to accurately appraise the findings of the trial within the context of the robustness of its methodology. It would also allow accurate identification of trials with the highest level of evidence

to compare and combine their outcomes in systematic reviews and meta-analyses (Turpin et al 2005).

A number of studies, however, suggested that the introduction of CONSORT did not improve reporting of RCTs (Hill et al 2002, Moher et al 2001, Montori et al 2002, Mills et al 2005). These findings seem to be consistent and seem to be shown over a period of time after the publication of CONSORT. The findings come in tandem with the fact that an increasing number of journals declare their endorsement of the CONSORT statement (Moher et al 2001). Mills et al (2005) examined 200 RCTs published in 2003 in five general and 10 specialty journals endorsing CONSORT in order to determine the extent to which these journals implement their recommendations. This study concluded that those journals, although adopted CONSORT, did not seem to enforce it consistently.

In the attempt to explore the impact of CONSORT on reporting the trials included in this review, the focus was on the issue of the lack of information. There is, albeit not significant, evidence that trials published after 1996 seem to have more information on quality criteria and a larger number of positively scored items compared with trials published earlier which seems to point to some positive impact from CONSORT. Koes et al (2005) in their assessment of back pain RCTs found no evidence of improvement in trials quality between 1960 and 2004, however it is not clear how the ‘don’t know’ scores (lack of information) were analysed.

The lack of information can be considered a sign of poor actual conduct and ‘don’t know’ scores, therefore, merged with negative scores (Schulz et al 1995). This

method was tested in this analysis and was found to change judgment of the quality of some individual criteria across trials from positive to negative (or vice versa). The problem with this approach is the obvious risk of describing a well conducted but poorly reported trial as a poor quality trial which could invalidate a valid outcome.

5.5.4 The association between trials' quality and other characteristics

The majority of trials included in this review were on non-pharmacological treatments (91 trials, 77%), which reflects the role of these types of treatments in the primary care management of LBP (van Tulder 1997). One of the main findings is the distinctly high proportion of negative scores for blinding of care providers and patients for a large number of trials in the occupational healthcare and physiotherapy settings. Blinding of patients and care providers is considered, rightly, as an integral part of conducting a high quality RCTs to provide an important safeguard against potential bias (Montori et al 2002) and the explicit reporting of which has been called for in the CONSORT statement (Begg et al 1996). More importantly, perhaps, empirical evidence suggests that methodological flaws in trial design such as inadequate patient and care provider blinding are associated with overestimated treatment effect (Schulz wet al 1995). The fact that the majority of trials included in the review have inadequately blinded care providers and participants would understandably raise concerns about the validity of their outcomes.

However, it could be argued that this study's findings might be expected considering the known difficulty in adequately blinding participants and care providers in trials on non-pharmacological treatments (Black 1996, Boutron et al

2003, 2004), such as education, exercises programmes, manual therapy or psychological treatments. Perhaps a quality assessment checklist such as the one used here is not suitable for trials on non-pharmacological treatments. Trials on these types of treatments might require a particular way to assess their quality to take into account particular aspects of their methodology such as difficulty in blinding. The argument could be extended to involve trials on other treatments such as some complementary and alternative treatments.

The relevance of blinding, however, might also be related to the type of trial as well as the type of treatment. As mentioned earlier, it might be crucially relevant in an explanatory trial that aims to test the efficacy (specific effects) of a treatment (regardless of the type being pharmacological or non-pharmacological), compared with a pragmatic trial which aims to assess the overall effects of a treatment as used in clinical practice, including both its specific effect and the effects of non-specific factors.

For these reasons it could be argued that the high proportions of negative scores for blinding patient and care provider blinding were expected for the large proportion of non-pharmacological trials included. What was not expected and harder to explain, however, were the findings of high negative scores for criteria such as adequacy of random sequence generation, managing co-interventions and reporting drop-out rate in these same trials. There does not seem to be an obvious reason why trials should have unsatisfactory conduct in these important areas.

Trials on pharmacological treatments were more satisfactory in adequacy of random sequence generation, concealment of allocation to treatment arms, blinding and managing co-interventions. However, almost a third of these trials were still not satisfactory in controlling for co-interventions. Also, information was not available on each of the criteria in an average of a third of the trials. An interesting finding was the scale of lack of information on care provider blinding, which was missing in about six times the proportion of pharmacological trials as those on non-pharmacological treatments. The number of these types of trials, however, was small and therefore caution is required before generalising these results.

5.6 Limitations

The trials that were assessed for quality were included according to specific inclusion criteria in the original systematic review conducted to examine responses to primary care treatments. Many trials were excluded because they included either ineligible populations, treatments, outcomes or settings. However, the study satisfied the specific objectives. Findings in this thesis although provided important insight, might not be generalisable to all RCTs in back pain.

Assessors of trials' quality in this study were not blinded. Although some studies have suggested that blinded assessment would be associated with reduced reviewer and selection bias and provide a better consistency in the final quality scores (Jadad et al 1996) other studies did not find the same association (Verhagen et al 1998). Furthermore, blinding would not have been consistently

successful when the trials are from the field of interest and expertise of the reviewers.

5.7 Summary

Only about half of back pain trials included in this study were of high quality with a wide variation among trials. Information on important quality criteria was missing in many trials. Further to the previous repeated calls for authors to adhere to the CONSORT statement in reporting, this study shows that further efforts to address this situation are needed. The findings also confirm the importance of describing the extent of the missing information on quality criteria separately rather than simply adding them to the positive or negative scores.

It seems that a quality assessment tool with the flexibility that provides the potential to allow it to be used with differing types of trials (exploratory or pragmatic) or treatments (pharmacological or non-pharmacological) might go a long way in addressing some of the main issues outlined in this review, such as those around patient and care-provider blinding. It seems that the new Cochrane Risk of Bias tool might represent an example of such a tool.

Not all back pain trials were assessed for their quality in this study, but only RCTs that were included in the systematic review. It is, therefore, important not to generalise the findings to all back pain trials. Another objective was to examine the association between quality and responses to treatments and this is the subject of the next chapter, Chapter Six.

Chapter Six

Factors associated with responses to treatments for low back pain: a meta-regression analysis of randomised clinical trials

6.1 Introduction

In Chapter Four responses to a wide range of primary care treatments in clinical trials for non-specific low back pain were studied. Further to the common pattern these responses appeared to follow, the review also showed a wide variation in the size of those responses which is the focus of this chapter.

Variation in 'treatment effects' (heterogeneity) is common in systematic reviews (Higgins & Green 2008). It is defined as the amount in which the 'effect' in each trial deviates from the pooled 'effect' (Thompson & Higgins 2002). Exploring possible sources of heterogeneity is important (Egger et al 2009, Thompson 1994) as they would represent factors that would be associated with or influence 'treatment effects' (Thompson 2009). The aim in this chapter was to study possible sources of heterogeneity in overall responses to treatments.

The principles and features of heterogeneity and its assessment for treatment effects should be the same when investigating heterogeneity in overall symptom progression. A possible difference, and arguably an advantage of studying sources of heterogeneity in overall responses to treatments, is that these responses represent the effects of all factors, specific and non-specific, and therefore include all possible sources of heterogeneity between responses.

Clinical as well as methodological diversity among trials contribute to heterogeneity in responses to treatments. The type of treatments used (for example, pharmacological or non-pharmacological), intensity and duration of symptoms, patients' socio-demographic characteristics as well as quality and

setting of the trials are examples of trial aspects that might influence response treatment. These trial characteristics represent, by definition, non-specific factors hence the interest in this thesis in studying them and their association with responses to treatments. Further, in Chapter Five, the findings showed that many (over a half) of the RCTs included in the systematic review were of low quality and information on a number of quality criteria was missing. In this chapter, the potential of the influence of trials overall quality, as well as individual aspects of quality, on responses to treatments will be explored.

6.2 Objectives

The main objective was to evaluate the extent to which heterogeneity in the size of responses to treatments for NSLBP described in Chapter Four is explained by the variation in the following trial characteristics: 1) patients' sociodemographics; 2) pain characteristics; 3) trial setting and quality. The observation in Chapter Four that the effect of non-specific factors was unrelated to the type of treatment and could in fact be stronger than the effect of the type of treatment was also tested by examining the association between type of treatment and symptom progression.

6.3 Methods

6.3.1 Criteria for trials' inclusion

Data from randomised clinical trials included in the systematic review (Chapter Four, Methods section) were used.

6.3.2 Outcomes

The main outcome was overall symptom progression within each trial arm. Although data on symptom progression were collected for pain intensity (VAS for pain and its equivalents) and functional disability (RMDQ and ODI), pain intensity was the outcome measure that was used by the largest number of trials and was, therefore, the focus of analyses reported here. Similar analyses were performed for functional disability and reference to and comparison with the results for pain intensity will be made.

Symptom progression was represented by the standardised mean change (SMC) calculated as the difference between mean outcome at baseline and follow up divided by the standard deviation (SD) at baseline (Morris 2000). Where means and standard deviations were not provided, these were calculated from the available data (see Chapter Four). Three common follow up time points were used for analyses, 13, 27 and 52 weeks.

6.3.3 Characteristics

These were pre-selected trial characteristics that are commonly identified in the literature as potential non-specific factors that could be associated with responses to treatments in clinical trials (Evers et al 2009, Macedo et al 2008, Pitz et al 2005, Wasan et al 2010). They were related to participants (age and gender), symptoms (duration) and methods (trial setting and trial quality). The quality assessment criteria checklist of the Cochrane Back Review Group (van Tulder et al 2003) was used to assess the internal validity of the included trials as described in Chapter Five. To examine whether the size of overall responses to treatments varied according to the type of treatment, the association between treatment type and

responses was examined using the two types of classification of treatment types: index, active comparator, placebo, usual care or waiting list and pharmacological or non-pharmacological.

The available evidence on the influence of these characteristics or factors vary greatly and mainly comes from fields other than back pain, such as osteoarthritis (OA) (Zhang et al 2008), migraine headache (Evers et al 2009, Macedo et al 2008) and irritable bowel syndrome (IBS) (Pitz et al 2005, Dorn et al 2007). This also applies to the evidence related to the association between trials quality and symptom progressions, as was discussed in Chapter Five. Furthermore, in the majority of the studies, these factors were studied for their association with treatment effect rather than overall symptom progressions. The exploration of the association of these factors on responses to treatments in this study, therefore, was not based on any *a priori* assumptions or hypothesis related to any of the selected characteristics.

To facilitate the interpretation of analysis, data on some characteristics were dichotomised. Trials were dichotomised around the overall mean age for the included trials' populations of 43 years and the mean percentage of females of 52%. For pain duration, trials were dichotomised as 'chronic' pain trials and 'acute/subacute' trials and the latter group was referred to as acute. This relied on the description provided by authors and inclusion criteria. This particular distinction between subacute and chronic (usually around a cut-off point of 12 weeks, which is an accepted minimum duration for chronic pain), was used to avoid the potential problems associated with the diversity in defining 'subacute' back pain in trials (Pengel et al 2002).

Two types of classifications were used to subgroup trial arms according to type of treatment. The first was whether they were placebo, index, active comparator, usual care or waiting list (i.e. five groups). The second classification was according to whether the treatments were pharmacological or non-pharmacological.

Trials' settings were split into general practice, occupational healthcare units, physiotherapy departments, general population, mixed setting and other in which the setting was not clear (i.e. six groups). Trial quality involved assessment of each individual quality criteria ('yes', 'no' and 'not clear'). Results of these assessments were then summarised for each trial and the overall summary score dichotomised into high or low quality, defined according to at least 50% of the quality criteria being satisfied, i.e. five or more positive scores, as described in detail in Chapter Five. Studies with the 'not clear' category for individual quality criteria were dropped from analyses for those individual criteria to follow the argument for not adding 'don't know' to the negative scores. Also, there was no logical justification to study the association between symptom progression and the 'not clear' scores, which are related to reporting of the trial rather than the conduct of the trial itself.

6.3.4 Analysis

Meta-regression analyses (Thompson & Higgins 2002) were conducted to explore whether selected trial characteristics were associated with within-arm symptom progression. All arms of all included trials were included for the main analysis. Sensitivity analyses were conducted to evaluate the association in one randomly

selected arm from each trial, to overcome any potential cluster effect of within-arm responses in trials.

A random effects model meta-regression analysis was used (Lau 1998, van Houwelingen et al 2002) using STATA IC11 software. The significance level was set at 5% ($p = 0.05$) to test the significance of the association between determinant and outcome. The weight for each trial arm was equal to the inverse of its variance. Larger size trial arms, which have smaller variance, were hence given larger weight than smaller arms. The measure of variability between responses in trial arms (squared tau) was calculated.

The concept of random effects model is not exclusive to meta-regression analysis, but applicable to meta-analysis in a broader sense, as was explained in Chapter Four, Analysis section. To explain the concept briefly, a random-effects model assumes that there is variation between trial arms in the association of characteristic and outcome. Hence this analysis leads to an estimate of the "average" association between characteristic and outcome across trial arms (Lau et al 1998).

First, univariable analyses were performed to examine the unadjusted association between each of the selected trial characteristics and the SMCs. Multivariable analyses were then conducted to adjust for potential confounding. For these analyses pain intensity at baseline was selected as a potential confounder for the influence of pain duration as well as for the influence of gender.

The association of pain duration with symptom progression was expected to be potentially confounded by baseline pain severity. In other words, patients with longer duration LBP might have higher (or lower) pain severity than those of shorter LBP duration, which may explain part of the association between pain duration and outcome. Such a possible confounding was therefore explored. There is some evidence from the literature that women are more likely to report more severe pain and baseline pain intensity was therefore explored as a potential confounder for the association between gender and symptom progressions. Type of treatment was also assessed as a potential confounder for the association between response and some individual quality criteria that would be relevant to the type of treatment, such as patient blinding, practitioner blinding, drop out, concealment of allocation and compliance. However, only associations that were found to be significant in the unadjusted analyses were adjusted for potential confounding.

6.4 Results

Forty-four trials (Appendix 6) provided data for pain intensity outcome for 58 arms at 13 weeks, 59 arms at 27 weeks and 48 arms at 52 weeks. Thirty five trials provided data on the RMDQ, for 44 arms at 13 weeks, 24 at 27 weeks and 22 at 52 weeks. Twenty six trials provided data on the ODI, for 24 arms at 13 weeks, 20 at 27 weeks and 24 at 52 weeks. The results will be presented primarily for pain intensity outcome. The results on the RMDQ and the ODI will be presented separately and comparison made with pain.

6.4.1 Trial characteristics and symptom progression: Univariable analyses

Table 6.1 shows the unadjusted associations of trial characteristics with symptom progression for pain intensity.

6.4.1.a Socio-demographic variables: age, gender

Data on age were not available in 11 arms. 53% of trial arms included participants younger than 43 years old. The association with age was only statistically significant at 13 weeks, showing that trial arms with mean age younger than 43 years old was associated with larger symptom progression compared with mean age more than 43 years. Mean difference in pooled SMC was 0.37 (95% CI 0.05, 0.68).

Of the 44 trials providing data on pain intensity, three did not provide data on gender (eight arms). The mean percentage of females among trial arms providing data on pain intensity was 52%, which was not different from the mean for all the included trials (53%). There was a statistically significant association between the percentage of females in the trial arms and symptom progression at 27 and 52 weeks. Trial arms with lower percentage of females showed larger responses: at 27 weeks the difference in pooled SMC was 0.31 (95% CI 0.02, 0.59) and at 52 weeks, 0.38 (95% CI 0.05, 0.70). The association was not significant at 13 weeks but the size and direction were similar.

6.4.1.b Duration of pain

40% of trial arms included participants of whom the majority had an acute back pain episode at inclusion. At 27 weeks, arms from trials of participants with mainly

Table 6.1 Results of univariable meta-regression analyses showing mean difference in pooled SMC between trial arms for the selected characteristics for pain intensity, mean (95% confidence interval).

	13 weeks			27 weeks			52 weeks		
Trial characteristic	**n**	**Coeff (95% CI)**	**Tau²** 0.30	**n**	**Coeff (95% CI)**	**Tau²** 0.26	**n**	**Coeff (95% CI)**	**Tau²** 0.28
Mean age, *<43 years*	38:24	0.37 (0.05, 0.68)	0.28	32:22	0.19 (-0.10, 0.49)	0.24	28:19	0.26 (-0.08, 0.60)	0.27
Female, *<52%*	32:31	0.30 (-0.02, 0.61)	0.29	28:27	0.31 (0.02, 0.59)	0.23	18:34	0.38 (0.05, 0.70)	0.27
Pain duration, *acute*	34:33	0.20 (-0.10, 0.50)	0.29	30:29	0.39 (0.12, 0.65)	0.22	20:35	0.49 (0.19, 0.78)	0.23
Treatment type *(reference: usual care)*	8		0.28	7		0.26	7		0.29
Index	29	-0.38 (-0.87, 0.11)		25	-0.35 (-0.81, 0.11)		23	-.039 (-0.90, 0.11)	
Active comparator	25	-0.39 (-0.89, 0.11)		23	-0.34 (-0.81, 0.12)		21	-0.34 (-0.85, 0.17)	
Waiting list	3	0.06 (-0.81, 0.92)		4	-0.17 (-0.87, 0.52)		3	-0.56 (-1.37, 0.25)	
Placebo	2	0.31 (-0.63, 1.26)		0	-		1	-0.35 (-1.62, 0.92)	
Treatment type, *Non-pharmacological*	53:4	0.24 (-0.37, 0.86)	0.29	49:2	0.48 (-0.25, 1.22)	0.23	46:0	-	-
Setting (*reference: general practice*)	19		0.31	15		0.25	15		0.27
Occupational health care	18	-0.08 (-0.50, 0.33)		16	0.-22 (-0.61, 0.16)		13	-0.35 (-0.78, 0.08)	
Physiotherapy departments	8	0.21 (-0.34, 0.76)		6	0.08 (-0.45, 0.60)		4	0.17 (-0.49, 0.83)	
General population	9	-0.07 (-0.46, 0.60)		11	-0.35 (-0.77, 0.07)		5	-0.21 (-0.79, 0.36)	
Mixed setting	11	-0.12 (-0.59, 0.37)		11	-0.36 (-0.78, 0.06)		18	-0.34 (-0.74, 0.05)	
Other	2	0.48 (-0.44, 1.40)		0	-		0	-	
Quality: Overall trial quality, *low*	25:34	-0.02 (-0.33, 0.30)	0.31	28:27	-0.25 (-0.53, 0.03)	0.24		-0.45 (-0.73, -0.16)	0.23
Individual quality criteria: *not adequate*									
Randomisation	9:53	0.27 (-0.22, 0.76)	0.32	7:46	-0.33 (-0.79, 0.14)	0.27	5:33	-0.43 (-0.99, 0.14)	0.27
Concealment of allocation to treatment arms	0:41	-		0:37	-		2:24	-0.70 (-1.60, 0.20)	0.31
Patient blinding	54:13	-0.35 (-0.72, 0.02)	0.27	48:7	-0.43 (-0.85, 0.00)	0.24	46:9	-0.64 (-1.02, -0.26)	0.23
Care provider blinding	58:7	-0.05 (-0.54, 0.45)	0.31	54:5	0.05 (-0.46, 0.55)	0.26	46:9	-0.09 (-0.51, 0.33)	0.29
Co-interventions prevented	37:9	-0.07 (-0.60, 0.45)	0.38	27:9	-0.01 (-0.47, 0.44)	0.30	25:6	-0.08 (-0.61, 0.44)	0.27
Co-interventions standardised	31:6	0.08 (-0.61, 0.76)	0.47	25:2	0.63 (-0.25, 1.50)	0.31	25:0	-	-
Compliance	5:29	-0.56 (-1.13, 0.02)	0.30	2:31	-0.68 (-1.24, -0.12)	0.28	7:24	-0.68 (-1.16, -0.20)	0.25
Drop out	25:42	-0.02 (-0.33, 0.30)	0.31	19:40	-0.01 (-0.31, 0.30)	0.26	17:36	0.01 (-0.32, 0.35)	0.27
Measurement comparable**	2:65	0.34 (-0.56, 1.23)	0.30	0:59	-		0:55	-	
Intention to treat analysis	8:50	0.03 (-0.48, 0.55)	0.33	6:42	-0.46 (-0.93, 0.01)	0.23	2:39	-0.63 (-1.40, 0.13)	0.25

Pooled SMC at 13 weeks: 0.99 (0.86, 1.11), 27 weeks:1.07 (0.93, 1.20) and 52 weeks: 0.91 (0.76, 1.05), Coeff = coefficient from meta-regression. Positive (negative) coefficient indicates larger (smaller) response between baseline and follow-up compared to reference group. Underlined indicates significant (p<0.05). n: represent number of arms in the selected group compared with reference group. **Timing of measurement of outcomes comparable across groups.

acute back pain episode showed a significantly higher pooled response estimate compared with those with chronic episodes. However, this was statistically significant only at 27 and 52 weeks with mean difference in pooled SMCs of 0.39 (95% CI. 0.12, 0.65) and 0.49 (95% CI 0.19, 0.78) at these times respectively. This suggests that symptom progression was larger among acute and subacute LBP patients compared with chronic LBP patients. At 13 weeks, the association was not statistically significant. However, the direction of difference at this time was the same. It is also noted that the size of the difference in response increased with follow up time.

6.4.1.c Type of treatment

Using both classifications for type of treatment, there was no significant association between types of treatments and symptom progressions at any follow up times. However, there were very small numbers of arms using placebo and pharmacological treatments. Caution is therefore warranted interpreting the related results.

6.4.1.d Trial methods: Setting and quality

There was no significant association between setting of trials and symptom progressions at any follow up time.

Of the 44 included trials, 20 (46%) were considered high quality trials using the quality overall summary score (Chapter Five). Trial quality was associated with symptom progression, with low quality trials being associated with smaller within-arm responses to treatments, but this was statistically significant only at 52 weeks (mean difference of pooled SMC was -0.45 (95% CI -0.73, -0.16). At 13 & 27

weeks, the associations were not significant. However, the size of the difference in response between high and low quality trials was smaller but the direction was similar at 27 weeks to that at 52 weeks.

Of the individual quality criteria, the criteria of adequacy of patient blinding and compliance were statistically significantly associated with responses to treatments at 27 & 52 week follow up. When these two criteria were not satisfied, this was associated with smaller symptom progression. At 13 weeks, although these associations were not statistically significant, the size and direction of the associations were similar to those at 27 & 52 weeks. There was not sufficient number of arms to analyse the associations of the criteria related to concealment of allocation to treatments at 13 weeks, timing of outcome measurements at all follow up times and for the criterion of standardisation of co-interventions at 52 weeks. Apart from a small number of exceptions, in general the associations suggest that inadequate satisfaction of quality criteria was associated with smaller responses to treatments.

6.4.1.e Residual heterogeneity

At 13 week follow up, between-arm variance represented by squared tau without any independent variable included in the analysis was 0.30. Introducing the selected independent variables into the analyses did not further explain the variation to a large extent.

At 27 week follow up, squared tau without any independent variable included in the analysis was 0.26. The smallest residual variance was 0.22 following the inclusion of duration of pain in the model. This means that duration of pain

explained 15% of the between-arm variation in responses to treatment at 27 weeks. Variance was reduced to a lesser extent (0.23) when the gender was introduced, thus explaining 12% of heterogeneity.

At 52 weeks, between-arm residual variance without the introduction of any independent variable was 0.28. The largest reduction in between arms variation or heterogeneity was to 0.23, achieved with the introduction of duration of pain, overall trial quality and adequacy of patient blinding. This indicates that each of these determinants explained 18% of the variation in responses at this time point. Squared tau was reduced to a lesser extent with the introduction of compliance, to 0.25, i.e. compliance explained 11% of variation in responses at 52 weeks.

6.4.2 Multi-variable analyses (Table 6.2)

4.6.2.a Influence of gender adjusted for baseline pain intensity

Adjusting the association between gender and symptom progression for pain intensity at baseline resulted in similar estimates for gender as for the unadjusted analysis which were statistically significant at 27 and 52 weeks. Squared tau, reflecting the amount of variation in responses left unexplained, did not change at 13 weeks and only marginally fell at 27 weeks from 0.23 to 0.22 and at 52 weeks from 0.27 to 0.24. This suggests that the association between gender and symptom progression was not confounded by differences in pain severity at baseline.

Table 6.2 Mean difference in SMC between trial arms for the selected characteristics for pain intensity, adjusted for selected potential confounders.

Subgroups of trial characteristics	13 week Unadjusted	13 week Adjusted	13 week Tau^2	27week Unadjusted	27week Adjusted	27week Tau^2	52week Unadjusted	52week Adjusted	52week Tau^2
Model 1:									
Gender adjusted for baseline pain intensity									
Female < 52%	0.30 (- 0.02, 0.61)	0.29 (-0.02, 0.60)	0.29	0.31 (0.02, 0.59)	0.31 (0.04, 0.59)	0.22	0.38 (0.05, 0.70)	0.41 (0.09, 0.72)	0.24
Model 2:									
Pain duration adjusted for baseline pain intensity									
Pain type, <12w	0.20 (-0.10, 0.50)	0.17 (-0.13, 0.48)	0.28	0.39 (0.12, 0.65)	0.35 (0.08, 0.61)	0.22	0.48 (0.19, 0.78)	0.43 (0.14, 0.73)	0.22
Model 3:									
Patient blinding adjusted for type of treatment A [a]									
Patient blinding-Not adequate	-0.35 (-0.72, 0.02)	-0.26 (-0.67, 0.15)	0.27	-0.43 (-0.85, 0.00)	-0.37 (-0.83, 0.09)	0.25	-0.64 (-1.02, -0.26)	-0.60 (-1.01, -0.19)	0.24
Model 4:									
Patient blinding adjusted for type of treatment B [b]									
Patient blinding-Not adequate	-0.35 (-0.72, 0.02)	-0.33 (-0.72, 0.06)	0.27	-0.43 (-0.85,0.00)	-0.19 (-0.65, 0.26)	0.23	-	-	
Model 5:									
Compliance adjusted for type of treatment A [a]									
Compliance-Not adequate	-0.56 (-1.13,0.02)	-0.54(-1.16, 0.08)	0.28	-0.68 (-1.24, -0.12)	-0.59 (-1.16, -0.02)	0.27	-0.68 (-1.16, -0.20)	-0.58 (-1.07, -0.10)	0.24
Model 6:									
Compliance adjusted for type of treatment B [b]									
Compliance-Not adequate	-0.56 (-1.13, 0.02)	-0.63(-1.53, 0.28)	0.31	-0.68 (-1.24, -0.12)	-0.74 (-1.57, 0.08)	0.25	-	-	

[a]: according to classification of usual care, index, active comparator, waiting list and placebo. [b]: according to classification of pharmacological and non-pharmacological.
Positive (negative) coefficient indicates larger (smaller) response between baseline and follow-up compared to reference group. Highlighted indicates significant (p<0.05)

6.4.2.b Influence of duration of back pain episode adjusted for baseline pain intensity

Trials among chronic back pain patients had higher average pain intensity scores at baseline compared with trials among acute and sub-acute pain patients. Baseline pain intensity scores for chronic LBP trials were: mean 53, median 53, range: 19-79.7 and for acute LBP trials: mean 37, median 44.4, range: 20-61.

The association between pain duration and symptom progression remained significant and of similar magnitude after adjustment for baseline pain intensity. Squared tau fell only marginally at 13 weeks and 52 weeks, from 0.29 to 0.28 and from 0.23 to 0.22 respectively, and remained the same at 27 weeks.

6.4.2.c Influence of individual quality criteria adjusted for type of treatment

The association between symptom progression and the criteria of adequacy of patient blinding and compliance was adjusted for type of treatments. The number of pharmacological treatment arms was too small at 52 week follow up to allow for these analyses. The association between these two criteria and responses remained significant after adjusting for the type of treatments, except for patient blinding at 27 weeks.

The change in squared tau was marginal and not in a consistent direction after adjustment indicating insignificant effect of adjustment. The largest decrease in squared tau was at 13 weeks for compliance (from 0.30 to 0.28).

6.4.3 Functional disability

The results of univariable analyses with the RMDQ as the outcome are presented in Table 6.3. Sufficient data were not available to undertake the adjusted analyses.

Using the available data, pain duration (at all follow up times), gender (only at 52 weeks) and patient blinding (at 27 and 52 weeks) were statistically associated with responses to treatment on the RMDQ as they were for the pain intensity outcome. The differences between these two outcomes were for age, overall trial quality and compliance (not significantly associated with responses on the RMDQ), setting (significantly associated with responses to treatments on the RMDQ) and drop-out rate (significant for RMDQ at 13 weeks only). The associations with setting were significant at all follow up times for occupational health care and mixed setting trials, and only at 13 & 27 weeks for the general population setting trials. The direction in all these associations suggests that, in relation to the reference category which was general practice setting, they all were associated with smaller responses to treatments, with the largest difference in response being at 27 weeks for occupational healthcare setting (-1.54, 95% CI -2.35, -0.74).

There was a fewer trial arms that used ODI and provided data for the various types of treatments at 27 and 52 week follow up. At 27 weeks, no arm used placebo or usual care, 10 arms used index or active comparator treatments and two arms used waiting list control. At 52 weeks, no arms used placebo or waiting list control, 12 used index treatment, 13 active comparator treatment and two usual care. Because of the insufficient data, characteristics were not analysed in trials using this outcome.

Table 6.3 Results of univariable meta-regression analyses showing mean difference in pooled SMC between trial arms for the selected characteristics for RMDQ, mean (95% confidence interval)

Trial characteristic	n	13 weeks Coeff (95% CI)	Tau^2 0.35	n	27 weeks Coeff (95% CI)	Tau^2 0.43	n	52 weeks Coeff (95% CI)	Tau^2 0.39
Mean age, *<43 year*	23:22	0.06 (-0.29, 0.42)	0.32	13:17	-0.35 (-0.85, 0.16)	0.41	11:11	-0.04 (-0.64, 0.56)	0.41
Female, *<49%*	17:18	0.15 (-0.26, 0.57)	0.33	21:7	0.42 (-0.14, 0.98)	0.42	9:9	0.73 (0.15, 1.31)	0.28
Pain duration, *<12w*	18:27	0.75 (0.47, 1.03)	0.18	14:16	1.02 (0.69, 1.35)	0.17	10:12	0.70 (0.20, 1.20)	0.27
Treatment type *(ref: usual care)*	6		0.31	3		0.45	4		0.39
Index	20	0.05 (-0.50, 0.60)		12	-0.41 (-1.32, 0.50)		11	-0.31 (-1.12, 0.51)	
Active comparator	16	-0.01 (-0.58, 0.56)		15	-0.37 (-1.26, 0.53)		7	-0.56 (-1.43, 0.30)	
Waiting list	0	-		0	-		0	-	
Placebo	3	-0.56 (-1.39, 0.27)		0	-		0	-	
Treatment type, *Non-pharmacological*	34:4	0.22 (-0.38, 0.82)	0.30	25:0	-		14:0	-	
Setting *(reference: general practice)*	11		0.22	2		0.21	8		0.22
Occupational health care	9	-0.88 (-1.32, -0.43)		7	-1.54 (-2.35, -0.74)		2	-0.96 (-1.78, -0.15)	
Physiotherapy departments	0	-		2	-0.25 (-1.28, 0.79)		2	0.75 (-0.13, 1.62)	
General population	5	-0.92 (-1.47, -0.37)		7	-1.45 (-2.26, -0.64)		3	-0.92 (-1.63, -0.21)	
Mixed setting	9	-0.60 (-1.05, -0.14)		6	-1.42 (-2.25, -0.59)		5	-0.25 (-0.88, 0.39)	
Other	11	-0.57 (-0.99, -0.14)		6	-0.73 (-1.55, 0.09)		2	-0.32 (-1.16, 0.53)	
Quality: Overall trial quality, *low*	31:14	0.32 (-0.05, 0.70)	0.30	21:9	0.15 (-0.41, 0.71)	0.44	9:13	0.11 (-0.50, 0.71)	0.41
Individual quality criteria, *quality not adequate*									
Adequacy of randomisation	3:34	-0.43 (-1.20, 0.35)	0.32	1:29	-0.75 (-2.27, 0.77)	0.43	1:17	-0.81 (-2.38, 0.76)	0.42
Concealment of allocation to treatment arms	0:30	-	-	0:17	-	-	0:8	-	-
Patient blinding	28:13	-0.41 (-0.79, -0.03)	0.29	28:2	-1.23 (-2.14, -0.32)	0.34	16:6	-0.61 (-1.21, -0.02)	0.32
Care provider blinding	39:4	-0.13 (-0.75, 0.49)	0.32	30:0	-	-	20:2	-0.19 (-1.22, 0.83)	0.41
Co-interventions prevented	29:12	0.18 (-0.24, 0.61)	0.34	17:9	0.44 (-0.12, 1.00)	0.41	14:4	-0.06 (-0.85, 0.73)	0.40
Co-interventions standardised	20:6	0.44 (-0.13, 1.01)	0.32	17:0	-	-	14:0	-	-
Compliance	2:35	-0.30 (-1.21, 0.60)	0.34	2:22	-0.33 (-1.42, 0.75)	0.47	2:12	-0.01 (-1.10, 1.08)	0.39
Drop out	24:21	-0.43 (-0.76, -0.10)	0.26	10:20	-0.39 (-0.91, 0.14)	0.41	13:9	-0.01 (-0.61, 0.60)	0.41
Comparable measurements*	3:42	0.47 (-0.28, 1.22)	0.31	1:29	0.75 (-0.77, 2.27)	0.43	1:21	0.82 (-0.66, 2.30)	0.39
Intention to treat analysis	4:34	-0.54 (-1.17, 0.10)	0.34	2:26	-0.62 (-1.64, 0.41)	0.45	4:14	-0.70 (-1.50, 0.11)	0.42

Pooled SMC at 13 weeks: 0.79 (0.61, 0.97), 27 weeks: 0.97 (0.73, 1.21) and 52 weeks: 0.98 (0.73, 1.23). Coeff = coefficient from meta-regression. Positive (negative) coefficient indicates larger (smaller) response between baseline and follow-up compared to reference group. Underlined indicates significant (p<0.05). n: number of arms in the selected group compared with ref group.
*Timing of measurement of outcomes comparable across groups.

6.5 Discussion

The aim was to assess the extent to which trial characteristics explain the variation in responses to primary care treatments for non-specific low back pain in clinical trials.

The findings suggest that age, gender, pain episode duration and trial quality could explain some of the heterogeneity of responses. A large proportion of variation still remained unexplained suggesting that other factors are contributing to the variation in responses that were not examined here. For this analysis factors were preselected for which sufficient data were available in the included trials. An important finding was that the type of treatment did not seem to be responsible for variation in responses in the trials studied, although the numbers were small for some treatments e.g. pharmacological treatments. This is in line with the proposed hypothesis of a similar pattern of responses regardless of the treatments used.

Studies have been conducted to explore the association between certain trial characteristics and ‘effect’ of treatments for various medical conditions (Schulz et al 1995, Moher et al 1998, Kunz & Oxman 1998, Balk et al 2002) including low back pain (Wasan et al 2010).The study in this thesis is different in that the overall responses to treatments were studied rather than treatment effect. This would provide the opportunity to study factors that are usually controlled for and neutralised when treatment effects are studied.

The findings will be discussed in detail below and compared with evidence from the literature.

6.5.1 Heterogeneity in systematic reviews

A degree of variation in treatment effects or responses to treatments is expected in systematic reviews of clinical trials. In fact, some even argue that it is unreasonable to expect complete homogeneity among a group of trials in any systematic review (Hardy & Thompson 1998). Trials are likely to differ in population type, their socio-demographic characteristics, setting, types of treatments and their dosage, nature and severity of symptoms and methods of symptom assessment or outcome measures and trial protocol and conduct.

6.5.2 Sources of variation in symptom progression

6.5.2.a Gender

Trial arms with larger percentages of females showed generally smaller size of response. The association between gender and musculoskeletal pain has been explored (Elliott et al 1999, Smith et al 1999, Wijnhoven et al 2006 & 2007). In these studies, women were found to be more likely to report pain and consult their general practitioners with pain than men and more likely to report higher pain severity and higher expressed needs. These differences, however, were not always statistically significant and varied according to site of pain. Furthermore, the implication of these results on responses to treatments and how these responses differ in the two gender groups is not clear. Observational studies have explored gender as a possible prognostic factor (Hestbaek et al 2003) however no conclusive evidence was found for its association with the course of back pain. Chenot et al (2008) conducted a secondary analysis of data of a cohort of 1342

(58% females) participants of an RCT and specifically examined sex differences in presentation, course and management of low back pain. They used the Hannover functional ability questionnaire to assess what they described as functional capacity. They concluded that women in their cohort tended to have lower functional capacity at baseline and after 12 months, however the differences between the gender groups were small and not statistically significant (scores for women at baseline 65 (95% CI 63.5, 66.5) and at 12 months 69.3 (95% CI 67.6, 71.1) vs men at baseline 70.7 (95% CI 68.9, 72.6) and at 12 months 78.2 (95% CI 76.3, 80.1)), p=0.886.

Evidence for the association between gender and treatment effect for back pain comes from a study conducted by George et al (2006) who concluded that for patients with acute low back pain, there was no difference in pain and disability outcomes between males and females assessed four weeks after starting physical therapy treatments. I am not aware of a study that assessed the direct association between gender and symptom progression for NSLBP.

6.5.2.b Duration of LBP

The finding that trial arms with patients with shorter duration back pain show larger symptom progressions than patients with longer duration was not unexpected. The historical classification of back pain according to its duration remains in use in clinical guidelines on management options, indicating varying prognosis depending on its duration. Observational studies also showed the important role of pain duration in the clinical course of back pain (Hestbaek et al 2003, Pengel et al 2003). These studies show that patients with shorter duration,

more recent or acute LBP, are more likely to show improvement in symptoms compared with patients with longer duration or chronic LBP.

6.5.2.c Type of treatment

The finding that responses in trial arms were not associated with type of treatment (index, active comparator, placebo, usual care, waiting list and pharmacological or non-pharmacological) confirms the observation in the systematic review presented in Chapter Four that the pattern of responses in trial arms was similar regardless of the type of treatment.

However, the number of trial arms in some groups was small which might have compromised the ability to detect any association between types of treatment and response. Also, treatments were classified according to general types rather than to particular types of treatments such as acupuncture, manual therapies, NSAIDs or exercise. As within-arm responses studied here represent both the specific effects of the treatment and the context effect, variation between the overall responses might have been explained, partly, by effects of specific types of treatments such as those mentioned above, however small or modest they might be. However, two important points needs to be stated here to address this issue. Firstly the objectives of the study were not intended to prove the lack of specific effect of treatments for NSLBP. In fact one of the main intentions for this study and the whole thesis was the supposition that RCTs do not seem to be able to provide a clear evidence for the specific effects of treatments not necessarily because of the lack of such an effect but because of the large size of the context effect. Secondly, this study was not designed to study a particular treatment. As will be mentioned later in this chapter, and is pointed out in Chapter Six and

Chapter Seven, studying the effects of non-specific factors in groups of trials on particular treatments, adopting a similar approach to this thesis', would address the issue discussed here but with objectives that would be different from those of this thesis.

Beyond the significance of the associations with types of treatments, it is interesting to note that responses in usual care and waiting list arms were (albeit not significantly) larger than those in index and active comparator arms. This seems to be counterintuitive as the expectation would be that response to the index treatment should be larger as a result of the specific treatment effects. However, usual care and waiting list participants, although not provided with the trial treatment (no-treatment arm), are in fact free to use and have access to any active treatment available. Although the types of treatments used by participants in the 'usual care' arm are rarely recorded, some of these treatments might be similar in effect or even superior to some of the index treatments used in the trials (Somerville et al 2008). This might explain the findings related to the size of response in these treatment arms. However, this is an indirect comparison between different types of treatments, directed more towards specific effects of treatment, which of course is not the objective of this thesis.

6.5.2.d Trials' quality

It has been suggested that treatment 'effects' estimated in poor quality trials could be an under or over-estimation of the 'true' treatment effect and some studies have shown that trial's quality can influence treatment 'effect' size (Schulz et al 1995, Elliott et al 1999, Smith et al 1999), although this has been contradicted by others (Moher et al 1998, Kunz & Oxman 1998, Balk et al 2002, Verhagen et al

2001). I am not aware of any study that examined the association between quality and within-arm symptom progression in back pain trials.

It is important for any comparison between these findings and findings of other studies to be interpreted within the context of the quality assessment process that was used in this thesis which might be different from the assessment used in other studies. The evidence from this study for the association between overall trial quality and within-arm responses to treatment was limited. The problems and caveats associated with using the quality summary score approach were addressed and discussed in Chapter Five. It is therefore difficult to draw solid conclusions on the association between response and trial overall quality.

6.5.2.e Individual quality criteria

In this thesis, adequacy of patient blinding (for pain and RMDQ), compliance (for pain) and drop-out rate (for RMDQ) were found to be associated with responses to treatments. It is interesting to also note that although the association with the other criteria were not significant, the direction was similar. When these criteria were satisfied this was associated with larger within-arm responses to treatments.

Some studies have found that treatment 'effect' was larger when patient blinding, as well as other quality criteria, was inadequate (Schulz et al 1995, 1996, Moher et al 1998). The explanation for those findings was assumed that when participants knew whether they received active or 'inactive' treatment they responded accordingly and the difference between the two groups was large (large effect size on a trial level). However, this is different in this thesis where within-arm responses to treatments were studied. Adjusting patient blinding for

type of treatment did not change the significant association, suggesting that it was independent of the type of treatment.

The issue of patient blinding is about patient's 'knowing' (or not) what type of treatment he is receiving. To correctly interpret the influence of this 'knowing' on symptom progression, it would be important to understand patients expectations from that treatment, any prior preference to a particular treatment, beliefs regarding the treatment and previous use of or experience with the treatment. Data on these factors were not available from a sufficient number of included trials to explore their influence.

The majority of the trials included non-pharmacological treatments (e.g. exercise, manipulation, osteopathy, psychotherapy, heat wraps) in which it is widely argued that it is very difficult and might be impossible to successfully blind patients or practitioners (Black 1996, Boutron et al 2003, 2004). However it has been recommended that even if blinding is not feasible, inadequate blinding can be a source of potential bias in a trial (van Tulder 2003). In the results presented in Chapter Five, patient blinding was judged adequate in 80% of trials on pharmacological interventions and in 12% of non-pharmacological trials. This indicates that in the majority of trials included in the review the treatments were mostly active, non-pharmacological and blinding was mostly inadequate and factors such as expectations, preference and previous use of treatment may have influenced responses in these arms.

Adequacy of compliance was also associated with larger symptom progression. Adjusted for the type of treatment, the significance of the association did not

change suggesting that better compliance was associated with larger responses regardless of the type of treatment used. Again the attention needs to be drawn to the large difference in the number of arms in the comparison groups. As was mentioned in the previous chapter, information on this criterion was available in 54% of the included trials and in the majority of the trials (45% of all included trials) this was satisfied. Assuming this is sufficient evidence, it might reflect a characteristic of compliant individuals showing larger symptom progressions. It might reflect their expectation regarding the treatment or their satisfaction with the treatment. On the other hand, compliance might in fact be a reflection of the participants experiencing (or not) benefit from the treatment i.e. those who experience benefit (large symptom progression) would of course continue using the treatment (positive compliance) and those who did not experience benefit (small symptom progression) would discontinue (poor compliance). Completing the missing information in trials published reports and having larger number of arms in the groups might help clarify some of these issues. Alternatively, it might also mean that better compliance actually leads to better (specific) effects of treatment.

6.6 Limitations

Some of the characteristics analysed in this chapter can be described as “true” trial characteristics (Schmid et al 2004) that apply equally to all participants within a trial or within a trial arm with no differences between participants, such as type of treatment or trial quality. Other characteristics although provided at trial level are in fact properties of individual participants and have different values among

participants, and are aggregated and presented at a trial level as a summary mean of the individual values. These patient level characteristics, such as age, gender, severity or duration of pain, are therefore not “true” trial characteristics. One caveat associated with using aggregated data is the lack of power to detect association of patients’ characteristics with outcome. What does this mean is that when a significant association is found through this analysis it is likely that it is an important and underestimated association. However, the fact that the findings might be biased is important to consider, as outlined below.

The use of aggregated data may yield biased results when associations are detected that are not present within the individual study samples, or actual associations are missed in the aggregated data (ecology fallacy or aggregation bias (Thompson & Higgins 2002). However, Schmid et al (2004), in a study to assess the utility of meta-regression in exploring the association between effects of treatment and trials’ and patients’ characteristics, concluded that in the absence of individual patients data, aggregated data could be used for this purpose with proviso that results are interpreted with caution and not generalised or applied to all patients.

It is important to pre-select or specify covariates or determinants based on a pre-analysis plan. Trials’ characteristics were preselected for this analysis to explore their association with responses to treatment. Due to the small number of trials providing information on some characteristics, multivariable analyses were restricted to include only two potential confounders and selected only a limited number of determinants to prevent data dredging.

The choice of determinants depended on the available data from the included trials. Determinants that would have been of interest and on which data were not widely available from the trials in the review include whether the treatment was provided in a public or private sector, the skills and expertise of the practitioners, whether the treatment was provided individually or in a group (e.g. exercises), patients' expectations and preferences to treatment.

The number of trial arms was sufficiently large for duration of pain, age, gender and trials overall quality. However, it was small for the type of treatments and some of the individual quality criteria.

To facilitate analysis, the continuous variables of age, percentage of females and duration of pain were dichotomised. It is suggested that the simplicity achieved through this is gained at the cost of loss of power and residual confounding (Royston et al 2006). It would be ideal for continuous variables to be entered in the analysis as continuous. However, it was judged for this thesis that the dichotomisation approach is adopted for the gain of being able to conduct more interpretable analyses.

6.7 Summary

There is evidence for significant associations between non-specific factors represented by trial characteristics and symptom progressions for non-specific low back pain such as age, gender, duration of pain, adequacy of patient blinding,

compliance and overall quality of the trial. The types of treatments did not contribute to the variation in responses.

These findings would explain some of the variation found in Chapter Four in the size of responses to treatments in clinical trials. Most of the variation remained unexplained which suggests that other unidentified non-specific factors may influence overall symptom progression.

To examine the association between responses to treatments and patient level characteristics, individual patient data (IPD) would be the basis for a more powerful analysis. This was conducted and the outcome is presented in Chapter Seven.

Chapter Seven

Factors associated with responses to treatments for low back pain: analysis of individual patient data (IPD)

7.1 Introduction

In Chapter Six trial characteristics were studied as possible source of heterogeneity in responses to treatments, using the aggregated data (AD). Some characteristics were characteristics of the participants (e.g. age, gender, duration of pain) aggregated at trial level. Using individual participant or patient data (IPD) analysis is the ideal method to confirm the association between such characteristics and responses to treatments (Lambert et al 2002). This was the first reason to carry out the analysis reported in this chapter.

The opportunity of using IPD was also taken to examine patient characteristics for which sufficient information was not available in the published trials included in the systematic review. Examples of these include patient expectation regarding helpfulness of treatment and patient preference for treatments. The importance of these characteristics, among other non-specific factors, was identified in a consensus study (Chapter Three) conducted among researchers who participated in a workshop during the International Low Back Pain Forum X meeting, held in Boston, USA in 2009. Patient expectations and preference for treatments were among 30 non-specific factors considered to be most influential on symptom progressions for low back pain (Box 3 Chapter Three).

Using IPD presents the potential for pooling data from multiple trials enabling more powerful statistical analyses. This is different from pooled analysis using AD as data at individual patient level are used to estimate the associations between non-specific factors and individual symptom progression within trials before pooled estimates are calculated.

IPD systematic reviews and meta-analyses were introduced in the nineties to satisfy the need for applying scientific evidence from clinical trials to individual patients. They are considered the gold standard source of data for collating and synthesizing evidence from clinical trials (Stewart & Clarke 1995, Simmonds et al 2005), superior to AD although the latter constitutes the basis for the vast majority of systematic reviews. The reason for the rarity of IPD systematic reviews and meta-analyses is that, compared with using AD, the whole process of using IPD, from obtaining the data to statistical analysis, requires more time and resources (Lambert et al 2002, Simmonds et al 2005, Riley et al 2010).

7.2 Objectives

The objective was to use individual patient data from randomised clinical trials to explore the influence on within-arm symptom progression of the following patient characteristics in terms of changes in pain and disability:

1. Age, gender, pain duration and history of back pain.
2. Expectations regarding helpfulness of treatment and preference for treatment.

7.3 Methods

7.3.1 Searching and selecting datasets

A systematic on-line search of Medline, EBSCO and the Cochrane CENTRAL database was conducted. This search was broader than the search conducted for the systematic review presented in Chapter Four, to enable the identification of

more recent trials and trials that could provide the data relevant for this IPD analysis.

7.3.1.a Criteria for inclusion

Included were randomised clinical trials among patients with non-specific low back pain published in the 10 years immediately prior to the year of the study, 2010.

7.3.1.b Participants and setting

Trials conducted among adult patients (aged 18 + years) consulting primary care services with non-specific low back pain that included data on the potential non-specific factors of outcome expectation and/or preference for treatments.

7.3.1.c Outcome measures

Outcome measures of interest were pain intensity (on VAS or similar, converted to 0-100 scale) and functional disability (RMDQ, ODI).

7.3.1.d Contacting authors

Clinical trials that met the inclusion criteria were identified and corresponding authors contacted, introducing the study and summarising the background, aims, objectives and the variables of interest. Authors who declared interest were sent the study protocol. Specific data request forms were completed and databases obtained either in SPSS, STATA or Excel format. Subsequent communications were made to clarify various issues in the obtained databases such as missing variables, definitions of variables and to clarify variable names and codes.

7.3.2 Analysis

Obtained databases were processed, prepared and analysed using SPSS 18 and STATA IC 11 software for Windows.

7.3.2.a Data preparation:

Relevant variables were identified and definitions and measurement scales checked. Variables were re-coded for uniformity across trials (e.g. gender: 0 male, 1 female). Some variables were dichotomised to facilitate analysis e.g. age was dichotomised around the mean age of all included trial populations.

7.3.2.b Factors influencing responses to treatments

Factors selected to assess their influence on symptom progressions included factors that had been shown to be of importance based on the analysis of AD presented in Chapter Six (i.e. patients socio-demographic characteristics e.g. age, gender, duration of current episode and treatment allocation). Also included were factors of patients' expectations regarding treatment and preference for treatment and history of back pain for which available AD was not sufficient to include in the meta-regression analysis.

For duration of current episode of back pain, data from all trials were dichotomised around 12 weeks. This is the commonly accepted cut off to distinguish between acute/subacute and chronic back pain and was adopted for the meta-regression analyses in the previous chapter. For history of back pain, participants were dichotomised into those with presence or absence of history of back pain.

7.3.2.c Expectations and preferences

Data on expectations regarding helpfulness of treatments were prepared to enable studying the influence of these characteristics on symptom progression. Expectation data were dichotomised into low and high expectation. Low expectation represented the low or negative half of the scale, e.g. 0-5 on 0-10 VAS scale, 0-2 on 0-5 scale and 'not helpful' on the other scales. High expectation represented the high or positive half of the scale, e.g. 6-10 on 0-10 VAS, 3-5 on 0-5 scale and 'expect helpful', 'very helpful' and 'yes' on the other scales.

A new variable for expectation was then created: among those who received a particular treatment, those who had low expectation for it *vs* those who had high expectation.

For the characteristic of preference for treatment, the influence of merely having a preference was studied. A new variable was created: those who did not have a preference (e.g. 'did not know', 'did not mind') versus those who did, regardless of the preference. A further new variable was created to study the influence on symptom progression of whether preference for a particular treatment was satisfied (by being allocated to the preferred treatment) or not. That is, among those who expressed a preference: those who were allocated to their preferred treatment *vs* those who were not. Those who did not have a preference were not included in this second variable.

7.3.2.d Examining the influence of factors on responses to treatments

The selected patient characteristics mentioned above were included in all analyses. Independent samples t-tests were performed to test the difference in

mean symptom progressions according to the selected non-specific factors. Associations were then adjusted for age, gender, duration of current episode of back pain and history of back pain using linear regression analyses.

In line with the main objective of this thesis of examining associations with symptom progression within-arm rather than with between-arm treatment effect, the influence of non-specific factors was analysed at the level of treatment arm. The interest in these associations was regardless of the type of treatment, i.e. index, active comparator or control treatment. Therefore all analyses were performed separately for each treatment arm in each trial.

7.3.2.e Pooling data from multiple trials

IPD from the included trials was assessed for comparability of the dependent and independent variables and follow up times to assess the feasibility of pooled analyses. RMDQ was the outcome most commonly used in the trials and was therefore used for the pooled analyses. The follow up time points of 13, 27 and 52 weeks were used as these were the most common follow-up periods across the trials and are similar to the time points used in the meta-regression analyses of the previous chapter.

Adopting a two-stage approach (Simmonds et al 2005), within-arm differences in symptom progressions (mean difference, 95% confidence intervals), according to the selected patient characteristics, were computed for each individual trial arm (stage 1, using adjusted analyses described above) and then pooled and analysed using standard meta-analysis methods described in Chapter Four (stage 2).

To assess the potential cluster effect of within-arm responses in each trial on the associations of factors with responses to treatments, sensitivity analyses were then performed. This was performed by extending stage 2 to a multivariable meta-regression analysis with the trials as an independent variable

7.4 Results

Thirteen trials met the inclusion criteria. Three authors declined the invitation to provide data and responses were not received from four authors. IPD was successfully obtained for six trials published in the period 1999 – 2007 (Table 7.1).

Trial sample size ranged from 186 in the Moffett et al trial to 1334 in the UK BEAM trial with arm sample sizes ranging from 80 for usual GP care in the Thomas et al trial, to 353 for manipulation in the UKBEAM trial. Three trials included two arms (Hay et al, Johnson et al and Thomas et al), one included three arms (Heymans) whereas the UK BEAM trial investigated four types of treatments and also used two types of setting (private and NHS) for one of its treatments (manipulation).

Table 7.1 Descriptive summary of trials included in IPD analyses

Trial	Population type / setting	Treatments	Age, year Mean (SD)	Female %	Pain baseline VAS, 0-100 Mean (SD)	RMDQ baseline Mean (SD)	Duration of LBP episode <1m (%)	History of LBP Yes (%)	Employment Yes (%)
Hay et al, 2005, UK	Primary care consulters with LBP, UK	-Manual physiotherapy -Brief pain management	41 (11.8)	48	55.6 (23.1)	13.53 (4.85)	74	74	50
Heymans et al, 2006, Netherlands	Workers sick-listed for LBP and who visited occupational physician, The Netherlands	-High intensity back school -Low intensity back school -Usual occupational therapy	40 (9.8)	21	65.8 (18.0)	8.62 (4.42)	33	81	100
Johnson et al, 2007, UK	Primary care consulters with LBP, UK	-Community-based treatment program, consisting of eight 2-hour group sessions over a 6-week period. Each group comprised between 4 and 10 participants and was led by 2 physiotherapists. -Back pain educational booklet and usual GP care.	48 (11.2)	60	48.3 (20.9)	10.76 (3.93)	46	-	77
Moffett et al, 1999, UK	Primary care consulters with LBP, UK	-Exercise program -Routine GP care	42 (8.9)	43	-	6.08 (4.00)	9	78	85
Thomas et al, 2006, UK	Primary care consulters with LBP, UK	-Acupuncture -Usual care [a]	42 (10.7)	61	-	32.93[b] (15.03)	-	84	79
UK BEAM team, 2004	Primary care consulters with LBP, UK	-Best GP care -Exercise -Manipulation -Combined exercise & manipulation	43 (11.2)	56	-	9.02 (4.02)	7	68	73

[a] Comprised a mixture of interventions including physiotherapy, manipulation, drugs and back exercises. [b]: ODI outcome

Age of participants ranged from a mean of 40 years (SD 9.8) in the Heymans et al trial, to 48 years (SD 11.2) in the Johnson et al trial, mean age across all trials was 43 years. Percentage of females ranged from 21% in the Heymans et al trial to 61% in the Thomas et al trial.

7.4.1 Symptom progression

Five trials used the RMDQ, three of these also used pain intensity outcome (VAS). One trial (Thomas et al.) used the ODI only. Mean responses to treatment on the RMDQ were calculated for the five trials within each trial arm (Appendix 7.1).

7.4.2 The influence of non-specific factors on symptom progression

Table 7.2 shows patients' characteristics that were selected to assess their influence on symptom progression. Data on age and gender were available from all trials; on history of LBP and duration of present episode of back pain from five trials; on expectation of helpfulness of treatment from four trials and on treatment preference from three trials.

The results of unadjusted and adjusted within-arms analyses of the associations between pre-selected characteristics and responses to treatments are outlined below. These findings will now be described in detail. For each selected characteristic, the results of within-arm analyses will be outlined followed by the results of the pooled analyses. Appendix 7.2 shows the results for the Thomas et al trial using ODI.

Table 7.2 Characteristics included in analyses to assess their influence on responses to treatment

UKBEAM	Hay et al	Heymans et al	Johnson et al	Moffett et al	Thomas et al
Age	Age	Age	Age	Age	Age
Gender	Gender	Gender	Gender	Gender	Gender
History of LBP	History of LBP	History of LBP		History of LBP	History of LBP
Duration of LBP	Duration of LBP	Duration of LBP	Duration of LBP	Duration of LBP	
			Presence of preference for treatment	Presence of preference for treatment	Presence of preference for treatment
			Satisfaction of preference for treatment		
Helpfulness of treatments		Helpfulness of treatments		Belief in effectiveness of index treatment	Expectation re-acupuncture

7.4.2.a Gender

Within-arm adjusted analyses showed that for pain and RMDQ (Table 7.3a, b) the evidence for the associations was weak and not consistent and the direction varied even within one trial (Heymans et al). The association was significant in three arms of two trials for RMDQ and only one arm for pain.

The results of the pooled analyses (Table 7.12) (13 arms, 2455 patients) did not show significant association between gender and symptom progressions for RMDQ for the selected follow up times. However, the direction of the associations (not statistically significant) suggests that men were more likely to have larger symptom progressions than women. The size of the difference in response ranged from 0.26 (95% CI -0.31, 0.84) at 13 weeks to 0.57 (95% CI -0.66, 1.81) at 27 weeks.

Table 7.3a Within-arm difference in the change in outcome scores (mean, 95% CI) according to gender (males compared with females) for pain intensity (0-100 VAS). Positive results indicate larger improvement in pain.

	13weeks			27weeks			52weeks		
	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted
Hay et al 2005									
-Manual physiotherapy	82:66	-3.53 (-12.50, 5.43)	-4.44 (-13.42, 4.55)				91:70	3.02 (-6.06, 12.09)	3.70 (-5.55, 12.87)
-Brief pain management	79:68	-3.84 (-13.45, 5.76)	-3.76 (-13.51,10.0)				88:72	1.81 (-7.99, 11.61)	2.16 (-7.74, 12.06)
Heymans et al 2006									
-Occupational health care	62:15	-3.0 (-14.2, 13.6)	-3.8 (-17.2, 9.6)	61:12	-0.9 (-17.9, 16.0)	-3.1 (-20.9, 14.8)	67:17	-1.9 (-16.6, 12.8)	-1.0 (-18.3, 14.3)
-Low intensity back school	52:15	-9.2 (-25.1, 6.8)	-8.5 (-24.7, 7.7)	49:16	-2.8 (-18.7, 13.0)	-2.4 (-17.6, 12.8)	64:19	-14.3 (-28.8, 0.2)	<u>-15.0 (-29.2, -0.8)</u>
-High intensity back school	50:18	7.0 (-3.2, 24.5)	10.2 (-0.41, 24.4)	45:17	-1.7 (-18.5, 15.1)	-3.1 (-20.3, 14.1)	52:16	-0.0.2 (-2.0.3, 1.9.9)	0.1.2 (-1.6.8, 1.9.3)
Johnson et al 2007									
-Physiotherapist led community-based treatment program	41:69	0.57 (-8.79, 9.93)	0.47 (-9.01, 9.95)	41:64	6.91 (-2.39, 16.02)	6.20 (-3.19, 15.59)	38:64	0.51 (-10.47, 11.49)	-0.11 (-11.23,11.44)
-Back pain educational booklet and usual GP care	47:66	1.63 (-7.86, 11.12)	1.38 (-8.24, 10.99)	40:58	10.36 (-0.84, 21.56)	10.40 (-0.86, 21.66)	37:57	9.56 (-1.60, 20.72)	9.41 (-1.78, 20.60)

n number of males versus females. Underlined indicates significant p<0.05

Table 7.3b Within-arm difference in the change in outcome scores (mean, 95% CI) according to gender (males compared with females) for RMDQ. Positive results indicate larger improvement in function.

	13weeks			27weeks			52weeks		
	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted
Hay et al 2005									
-Manual physiotherapy	89:73	-0.27 (-2.16, 1.61)	-0.44 (-2.20, 1.41)				93:72	0.55 (-1.35, 2.44)	0.77 (-1.13, 2.68)
-Brief pain management	33:74	0.53 (-1.57, 2.63)	0.49 (-1.62, 2.60)				88:76	1.96 (0.00, 3.91)	1.87 (-0.06, 3.81)
Heymans et al 2006									
-Occupational health care	62:15	1.97 (-3.91, 7.85)	1.28 (-4.52, 7.08)	59:12	2.24 (-3.80, 8.28)	2.03 (-3.86, 7.93)	67:17	4.25 (-0.69, 9.20)	4.74 (-0.36, 9.84)
-Low intensity back school	55:15	-6.19 (-10.75,-1.63)	<u>-5.56 (-10.04,-1.07)</u>	51:17	-3.55 (-8.28, 1.18)	-3.40 (-8.13, 1.33)	65:19	-3.56 (-8.05, 0.93)	-3.31 (-7.81, 1.19)
-High intensity back school	51:17	3.61 (-1.13, 8.35)	3.92 (-0.94, 8.77)	46:16	1.72 (-2.61, 6.05)	1.62 (-2.89, 6.14)	53:15	4.02 (-0.67, 8.72)	4.01 (-0.79, 8.81)
Johnson et al 2007									
-Physiotherapist led community-based treatment program	41:69	2.09 (0.31, 3.86)	<u>1.92 (0.11, 3.72)</u>	41:64	2.02 (0.29, 3.75)	<u>1.74 (0.00, 3.47)</u>	39:63	1.95 (-0.25, 4.14)	1.56 (-.68, 3.80)
-Back pain educational booklet and usual GP care	47:66	-1.09 (-2.84, 0.65)	-1.07 (-2.85, 0.70)	40:58	0.15 (-1.80, 2.11)	0.16 (-1.80, 2.13)	37:57	0.38 (-1.64, 2.40)	0.33 (-1.67, 2.34)
Moffett et al 1999									
-Exercise program	50:35	1.09 (-0.47, 2.64)	1.11 (-.47, 2.70)	46:31	1.99 (-0.14, 4.11)	1.97 (-0.19, 4.13)	50:33	1.97 (0.05, 3.88)	1.95 (-0.02, 3.91)
-Routine GP care	53:41	-0.14 (-1.71, 1.43)	-0.14 (-1.74, 1.46)	50:36	-.85 (-1.84, 1.13)	-1.04 (-3.03, 0.95)	51:37	-0.34 (-2.18, 1.49)	-0.36 (-222, 1.50)
UKBEAM 2004									
-Best GP care	114:142	0.64 (-0.44, 1.73)	0.61 (-0.51, 1.72)				105:142	-0.57 (-1.74, 0.61)	-0.57 (-1.80,0.66)
-Exercise	93:132	0.01 (-1.09, 1.12)	0.10 (-1.10, 1.30)				92:124	-0.46 (-1.63, 0.71)	-0.75 (-1.98, 0.47)
-Manipulation	108:179	0.18 (-0.97, 1.33)	0.11 (-1.16, 1.38)				104:169	-0.49 (-1.73, 0.75)	-0.60 (-1.96, 0.75)
-Combined exercise & manipulation	111:147	0.41 (-0.65, 1.46)	0.33 (-0.86, 1.52)				109:148	0.10 (-1.04, 1.24)	-0.26 (-1.51, 1.00)

n number of males versus females. Underlined indicates significant p< 0.05

The conclusion with regard to gender is that the evidence, weak and inconsistent as it is, seems to suggest a larger improvement in function (RMDQ) for men compared with women.

7.4.2.b Age

Within-arm adjusted analyses for both pain intensity and RMDQ (Table 7.4a&b) did not show evidence for a significant influence of age on symptom progressions apart from one treatment arm and at one follow up point (occupational healthcare arm, RMDQ at 27 weeks). The direction of the associations between age and responses to treatments was not consistent across treatment arms and trials.

When data for RMDQ from the five RCTs (13 arms, 2455 patients) were pooled, there was similarly no evidence for a significant influence of age on the pooled symptom progressions of included trials at the selected follow up times (Table 7.12) However, the direction of the non-significant differences in responses was the same in all follow up points in favour of younger patients suggesting that patients younger than 43 year old were more likely to show larger improvement in function (RMDQ) than older patients. The differences ranged from 0.17 (95% CI -0.28, 0.61) at 13 weeks to 0.45 (95% CI -0.80, 1.70) at 27 weeks.

Table 7.4a Within-arm difference in the change in outcome scores (mean, 95% CI) according to age (younger compared with older than 43 years) for pain intensity (0-100 VAS). Positive results indicate larger improvement in pain.

	13weeks			27weeks			52weeks		
	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted
Hay et al 2005									
-Manual physiotherapy	86:62	1.40 (-7.65, 10.45)	-0.51 (-9.79, 8.78)		-		92:69	-4.96 (-14.03, 4.10)	-4.83 (-14.19, 4.53)
-Brief pain management	80:67	-0.24 (-9.88, 9.40)	-0.22 (-9.94, 9.51)		-		88:72	-2.41 (-12.21, 7.39)	-2.22 (-12.09, 7.65)
Heymans et al 2006									
-Occupational health care	37:40	-2.3 (-13.3, 8.7)	-5.3 (-15.9, 5.4)	32:41	-1.0 (-13.7, 11.6)	-2.9 (-15.8, 10.1)	41:43	-0.3 (-12.1, 11.5)	-0.6 (-13.0, 11.8)
-Low intensity back school	39:28	2.9 (-10.7, 16.5)	1.5 (-12.9, 15.9)	35:30	9.1 (-4.5, 22.6)	8.5 (-4.7, 21.7)	45:38	7.8 (-4.5, 20.2)	8.3 (-3.9, 20.6)
-High intensity back school	39:29	-1.3 (-13.9, 11.3)	-1.6 (-14.8, 11.6)	33:29	-9.1 (-23.9, 5.8)	-11.3 (-27.4, 4.8)	37:31	1.8 (-12.8, 16.4)	1.0 (-14.6, 16.7)
Johnson et al 2007									
-Physiotherapist led community-based treatment program	41:69	-7.29 (-16.54, 1.97)	-7.29 (-16.67, 2.09)	40:65	-3.80 (-13.11, 5.52)	-3.86(-13.19,5.46)	37:65	-2.35 (-13.38, 8.69)	-2.57 (-13.86, 8.71)
-Back pain booklet & usual GP care	35:78	3.99 (-6.11, 14.08)	3.78 (-6.43, 13.99)	27:71	7.37 (-5.08, 19.81)	7.45 (-5.0, 19.88)	26:69	6.02 (-6.30, 18.33)	5.38 (-6.95, 17.71)

n number in the younger versus older groups

Table 7.4b Within-arm difference in the change in outcome scores (mean, 95% CI) according to age (younger compared with older than 43 years) for RMDQ. Positive results indicate larger improvement in function.

	13weeks			27weeks			52weeks		
	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted
Hay et al 2005									
-Manual physiotherapy	96:66	0.38 (-1.53, 2.28)	0.01 (-1.87, 2.07)		-		93:72	0.30 (-1.59, 2.19)	0.37 (-1.58, 2.32)
-Brief pain management	86:71	0.78 (-1.32, 2.88)	0.72 (-1.40, 2.84)		-		90:74	0.18 (-1.80, 2.15)	0.47 (-1.47, 2.42)
Heymans et al 2006									
-Occupational health care	37:40	3.06 (-1.56, 7.68)	2.77 (-1.83, 7.36)	32:39	5.36 (0.98, 9.74)	<u>5.01 (0.65, 9.37)</u>	42:42	3.07 (-0.92, 7.06)	3.37 (-0.75, 7.48)
-Low intensity back school	41:29	-4.97 (-8.78, -1.15)	-3.99 (-7.91, 0.07)	35:33	-2.49 (-6.61, 1.63)	-2.01 (-6.16, 2.14)	45:39	-3.39 (-7.14, 0.36)	-2.98 (-6.83,0.87)
-High intensity back school	37:31	1.81 (-2.36, 5.98)	2.26 (-2.12, 6.64)	32:30	-0.77 (-4.58, 3.04)	-0.58 (-4.74, 3.58)	36:32	-1.28 (-5.25, 2.69)	-1.29 (-5.40, 2.82)
Johnson et al 2007									
-Physiotherapist led community-based treatment program	41:69	-0.71 (-2.53, 1.10)	-0.78 (-2.56, 1.01)	40:65	-1.11 (-2.88, 0.66)	-1.18 (-2.90, 0.55)	36:65	-0.53 (-2.78, 1.72)	-0.74 (-2.98 1.50)
-Back pain educational booklet and usual GP care	34:78	0.56 (-1.32, 2.43)	0.54 (-1.35, 2.42)	27:71	1.29 (-0.85, 3.42)	1.30 (-0.87, 3.47)	26:68	1.16 (-1.03, 3.36)	0.94 (-1.27, 3.15)
Moffett et al 1999									
-Exercise program	45:40	0.40 (-1.15, 1.95)	0.42 (-1.17, 2.00)	41:36	0.90 (-1.23, 3.02)	0.73 (-1.43, 2.88)	42:41	0.11 (-1.82, 2.03)	-0.16 (-2.10, 1.69)
-Routine GP care	44:50	0.62 (-0.94, 2.17)	0.58 (-1.04, 2.19)	40:46	0.89 (-1.07, 2.86)	0.93 (-1.08, 2.94)	42:43	-0.11 (-1.92, 1.71)	0.04 (-1.84, 1.92)
UKBEAM 2004									
-Best GP care	112:130	0.93 (-0.15, 2.0)	0.65 (-0.46, 1.76)		-		111:123	1.05 (-0.10, 2.21)	0.86 (-0.36, 2.07)
-Exercise	91:125	-0.40 (-1.50, 0.69)	-0.44 (-1.63, 0.75)		-		81:125	-0.25 (-1.42, 0.93)	0.02 (-1.22, 1.26)
-Manipulation	127:158	-0.20 (-1.32, 0.92)	-0.31 (-1.55, 0.94)		-		114:157	-0.15 (-1.36, 1.06)	-0.13 (-1.47, 1.21)
-Combined exercise + manipulation	111:142	0.24 (-0.82, 1.29)	0.44 (-0.73, 1.61)		-		112:141	0.71 (-0.41, 1.84)	1.24 (-0.01, 2.47)

n number in the younger versus older groups. Underlined indicates significant p<0.05

7.4.2.c Duration of current episode of back pain

Measurements of duration of current episode of LBP varied across the five trials that provided data on this characteristic (Hay et al, Heymans et al, Johnson et al, Moffett et al & UKBEAM trials). The Hay et al trial was conducted among patients all with NSLBP of less than 12 weeks duration and therefore could not be included in this analysis.

For pain intensity (Table 7.5a), the adjusted within-arm analyses based on the two trials that provided data on this outcome (Heymans et al & Johnson et al trials) showed significant influence of duration of current episode of back pain on symptom progressions in two arms (both from the Heymans et al trial) only and at single follow up times (Occupational healthcare treatment arm at 13 weeks, and low intensity back school at 52 weeks). The direction of these significant associations, similar to that of the majority of the non-significant associations, suggests that shorter duration was associated with larger symptom progression.

For RMDQ (Table 7.5b), the adjusted association between duration and within-arm response was significant in three separate arms from two trials: occupational healthcare arm in the Heymans et al trial, at 13 weeks and exercise and combined treatment arms of the UK BEAM trial at 52 weeks. The direction of these significant associations, similar to that of the vast majority of the other non-significant associations, suggested that shorter duration was associated with larger responses to treatment.

Table 7.5a Within-arm difference in the change in outcome scores (mean, 95% CI) according to duration of current episode of back pain (12 weeks or shorter compared with longer than 12 weeks) for pain intensity (0-100 VAS). Positive results indicate larger improvement in pain.

	13week			27week			52week		
	n	**Unadjusted**	**Adjusted**	**n**	**Unadjusted**	**Adjusted**	**n**	**Unadjusted**	**Adjusted**
Heymans et al 2006									
-Occupational health care	63:13	16.9 (2.9, 30.9)	16.4 (2.3, 30.4)	61:11	4.3 (-13.2, 21.8)	4.9 (-13.3, 23.2)	71:12	-7.4 (-24.3, 9.5)	-7.2 (-24.6,10.2)
-Low intensity back school	60:7	17.1 (-4.5, 38.6)	17.0 (-5.4, 39.3)	56:9	22.6 (3.6, 41.6)	24.1 (5.0, 43.1)	74:9	19.7 (0.1, 39.2)	20.1 (0.9, 39.4)
-High intensity back school	60:8	3.8 (-15.5, 23.2)	2.7 (-17.7, 23.2)	52:10	6.7 (-13.7, 27.0)	13.1 (-8.9, 35.1)	56:11	9.0 (-10.9, 28.8)	11.8 (-9.8, 33.4)
Johnson et al 2007									
-Physiotherapist led community-based treatment program	72:38	-0.43 (-9.94,9.09)	0.13 (-9.55, 9.80)	69:36	3.97 (-5.56,13.50)	3.19 (-6.49,12.86)	68:34	1.27 (-9.99, 12.53)	1.63 (-10.11,13.37)
-Back pain educational booklet and usual GP care	64:49	3.14 (-6.29,12.56)	2.67 (-6.92,12.27)	53:45	0.76 (-10.48,12.0)	0.56 (-11.73,10.61)	51:43	5.92 (-5.13, 16.97)	4.88 (-6.21, 15.96)

n number of those with duration of 12 weeks or less versus longer duration back pain. Underlined indicates significant p<0.05

Table 7.5b Within-arm difference in the change in outcome scores (mean, 95% CI) according to duration of current episode of back pain (12 weeks or shorter compared with longer than 12 weeks) for RMDQ. Positive results indicate larger improvement in function.

	13week			27week			52week		
	n	**Unadjusted**	**Adjusted**	**n**	**Unadjusted**	**Adjusted**	**n**	**Unadjusted**	**Adjusted**
Heymans et al 2006									
-Occupational health care	63:13	8.40 (2.43, 14.37)	7.97 (1.90, 14.04)	60:10	6.60 (0.30, 12.90)	5.91 (-0.31,12.13)	71:12	2.05 (-3.74, 7.84)	0.93 (-4.84, 6.71)
-Low intensity back school	63:7	2.43 (-4.11, 8.97)	2.22 (-3.99, 8.43)	59:9	3.81 (-2.26, 9.88)	3.36 (-2.71, 9.43)	75:9	0.62 (-5.55, 6.78)	0.36 (-5.75, 6.46)
-High intensity back school	60:8	0.77 (-5.71, 7.25)	0.15 (-6.71, 7.01)	52:10	-0.47 (-5.65, 4.70)	-0.51 (-5.23, 5.21)	56:11	3.59 (-1.74, 8.93)	4.08 (-1.64, 9.81)
Johnson et al 2007									
-Physiotherapist led community-based treatment program	72:38	1.24 (-0.60, 3.08)	0.99 (-0.85, 2.83)	69:36	1.80 (0.01, 3.59)	1.59 (-0.20, 3.38)	67:34	1.13 (-0.18, 4.31)	1.84 (-0.48, 4.16)
-Back pain educational booklet and usual GP care	64:49	-0.16 (-1.91, 1.59)	-0.70 (-1.84, 1.70)	53:45	0.00 (-1.93, 1.93)	-0.14 (-2.09, 1.82)	51:43	1.60 (-3.53, 3.56)	1.47 (-0.52, 3.46)
Moffett et al 1999									
-Exercise program	53:32	0.29 (-1.31, 1.89)	0.35 (-1.26, 1.96)	48:29	0.60 (-1.59, 2.79)	-0.61 (-1.58, 2.79)	52:31	0.35 (-1.63, 2.33)	0.47 (-1.50, 2.44)
-Routine GP care	56:38	0.21 (-1.37, 1.80)	0.21 (-1.40, 1.83)	53:33	-1.60 (-3.59, 0.40)	-1.71 (-3.72,0.31)	53:35	-0.57 (-2.42, 1.27)	-0.61 (-2.49, 1.27)
UKBEAM 2004									
-Best GP care	101:144	1.24 (0.15, 2.34)	1.10 (-0.04, 2.23)				98:140	0.87 (-.35, 2.08)	0.71 (-0.54, 1.95)
-Exercise	92:125	1.24 (0.12, 2.36)	1.20 (-0.01, 2.41)				86:122	2.46 (1.29, 3.63)	2.06 (0.82, 3.29)
-Manipulation	113:161	0.61 (-0.57, 1.78)	0.42 (-0.84, 1.68)				102:157	0.72 (-0.55, 1.99)	0.37 (-0.98, 1.72)
-Combined exercise + manipulation	97:145	1.17 (0.06, 2.28)	1.08 (-0.21, 2.22)				94:147	1.51 (0.33, 2.70)	1.36 (0.06, 2.66)

n number of those with duration of 12 weeks or less versus longer duration back pain. Underlined indicates significant p<0.05

The results of the pooled analyses (Table 7.12) showed statistically significant influence of duration on symptom progression at 13 weeks (mean difference 0.76, 95% CI 0.25, 1.27) and at 52 weeks (1.03, 95% CI 0.49, 1.57). Although the association was not significant at 27 weeks its direction was similar.

The conclusion that could be drawn regarding duration of pain is that patients with back pain of 12 weeks or shorter duration were more likely to have larger improvement in function over time compared with patients with longer LBP duration.

7.4.2.d History of back pain

Information on history of back pain was available from five trials (only the Johnson et al trial did not provide such information). The scales used varied among trials (Table 7.6). In two trials (Hay et al & Heymans et al trials) information was provided on presence or absence of history of back pain (yes or no), in the other two trials (Thomas et al and UK BEAM trials) patients were asked to choose from the scale: 'none', '1-5' or '>5' episodes and in the Thomas et al trial, a choice of 'my pain never went away' was included in a similar scale. In the Moffett et al trial, patients were given an open question about the number of previous back pain episodes they had.

In spite of the variation in the scales, the data generally suggest that the vast majority of patients had a history of back pain (Table 7.6). In the three trials that provided data on the number of previous episodes (Thomas et al & UK BEAM trials), more than half of the patients had six or more previous episodes.

Table 7.6 Description of data on presence of history of back pain in included trials

Trial	Variable and scale		n (%)
Hay et al 2005	Previous history of LBP	Yes	296 (74)
		No	106 (26)
		Missing	0
Heymans et al 2006	Previous episodes of LBP	Yes	243 (81)
		No	46 (16)
		Missing	10 (3)
Moffett et al 1999	Number of previous episodes of LBP	Open question	Range 0-9, mean 2
Thomas et al 2006	Previous history of LBP	1=None	38 (16)
		2=1-5 episodes	80 (34)
		3=More than 5 episodes	121 (51)
		4=My pain never went away	0
UKBEAM 2004	Number of previous episodes of LBP	0=None	174 (13)
		1=1-5 episodes	346 (26)
		2=6-10 episodes	185 (14)
		3=More than 10 episodes	510 (38)
		Missing	119 (9)

Within-arm adjusted analyses for pain and RMDQ (Table 7.7a, b) showed that the association was significant only in one arm (best GP care, control arm, of the UK BEAM et al trial) for RMDQ and only at 52 week follow up. Patients of this trial who had no history of back pain showed smaller response to the best GP care than those who had history of back pain (mean difference in response on RMDQ 1.92 (95% CI 0.19, 3.66)). The direction of the associations was inconsistent in the other arms for pain and RMDQ.

The results of the pooled analyses (Table 7.12) showed significant association between presence of history of back pain and symptom progression only at 52 weeks. Absence of history of back pain was associated with smaller symptom progression for RMDQ at this follow up time.

Table 7.7a Within-arm difference in the change in outcome scores (mean, 95% CI) according to the presence of a history of back pain (presence compared with absence of a history of back pain) for pain intensity (0-100 VAS). Positive results indicate larger improvement in pain.

	13week			27week			52week		
	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted
Hay et al 2005									
-Manual physiotherapy	49:99	8.11 (-1.29, 17.50)	8.76 (-1.03, 18.55)				50:111	-2.50 (-12.22, 7.23)	-1.93 (-12.03, 8.18)
-Brief pain management	32:115	-2.05 (-13.67, 9.58)	-1.76 (-13.51, 1.00)				34:126	-3.68 (-15.59, 8.23)	-3.76 (-15.82, 8.31)
Heymans et al 2006									
-Occupational health care	11:65	14.9 (-0.3, 30.0)	14.7 (-0.4, 29.9)	9:63	9.2 (-9.7, 28.1)	9.9 (-9.7, 29.6)	10:73	0.4 (-17.9, 18.7)	1.0 (-16.3, 19.9)
-Low intensity back school	17:50	3.8 (-11.6, 19.1)	4.8 (-11.5, 21.0)	14:51	9.0 (-7.5, 25.6)	10.1 (-6.1, 26.2)	18:65	3.9 (-11.2, 19.0)	3.6 (-11.2, 18.4)
-High intensity back school	9:59	-7.7 (-26.0, 10.6)	-6.7 (-26.0, 12.6)	10:52	6.5 (-13.9, 26.8)	6.6 (-15.1, 28.3)	9:59	7.2 (-14.2, 28.7)	11.8 (-9.8, 33.4)

n number in the group with no history of back pain versus the group with history of back pain.

Table 7.7b Within-arm difference in the change in outcome scores (mean, 95% CI) according to the presence of a history of back pain (presence compared with absence of a history of back pain) for RMDQ. Positive results indicate larger improvement in pain.

	13week			27week			52week		
	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted	n	Unadjusted	Adjusted
Hay et al 2005									
-Manual physiotherapy	52:110	1.19 (-0.81, 3.19)	1.24 (-0.86, 3.33)				51:114	-1.29 (-3.31, 0.73)	-1.19 (-3.33, 0.95)
-Brief pain management	32:125	1.07 (-1.52, 3.67)	0.96 (-1.66, 3.58)				34:130	0.55 (-1.88, 2.98)	0.86 (-1.56, 3.28)
Heymans et al 2006									
-Occupational health care	11:65	3.58 (-3.09, 1-.24)	2.19 (-4.35, 8.73)	8:62	3.86 (-3.23, 10.95)	3.10 (-3.86, 9.97)	10:73	2.57 (-3.88, 8.82)	1.22 (-5.05, 7.48)
-Low intensity back school	17:53	-3.58 (-8.09, 0.94)	-2.25 (-6.75,2.24)	15:53	-2.50 (-7.48, 2.48)	-1.89 (-6.91, 3.13)	18:66	-2.12 (-6.74, 2.51)	-1.37 (-6.05, 3.32)
-High intensity back school	8:60	-0.34 (-6.82, 6.14)	0.45 (-6.41, 7.31)	9:53	-0.34 (-5.78, 5.07)	-0.57 (-6.49, 5.36)	8:60	-0.54 (-6.71, 5.63)	1.16 (-5.41, 7.72)
Moffett et al 1999									
-Exercise program	15:70	-0.10 (-2.13, 1.94)	-0.29 (-2.38, 1.79)	14:63	-0.81 (-1.94, 3.57)	0.53 (-2.25, 3.31)	14:69	1.04 (-1.52, 3.59)	0.81 (-1.79, 3.40)
-Routine GP care	8:86	-0.55 (-3.34, 2.24)	-0.35 (-3.24, 2.54)	8:78	-0.49 (-3.88, 2.89)	-0.22 (-3.67, 3.23)	8:80	1.14 (-2.0, 4.27)	1.14 (-2.13, 4.41)
UKBEAM 2004									
-Best GP care	35:201	-0.59 (-2.14, 0.96)	-0.66 (-2.20, 0.87)				33:197	-2.02 (-3.71, -0.33)	<u>-1.92 (-3.66, -0.19)</u>
-Exercise	31:179	-0.70 (-2.28, 0.88)	-0.46 (-2.14, 1.23)				30:171	-1.59 (3.24, 0.06)	-1.41 (-3.11, 0.29)
-Manipulation	29:229	0.27 (-1.63, 2.17)	0.28 (-1.66, 2.21)				25:215	-0.14 (-2.25, 1.97)	-0.19 (-2.33, 1.94)
-Combined exercise + manipulation	24:210	-1.73 (-3.57, 0.12)	-1.45 (-3.39, 0.50)				31:202	-2.11 (-3.88, -0.34)	-1.82 (-3.66, 0.03)

n number in the group with no history of back pain versus the group with history of back pain. Underlined indicates significant (p<0.05).

7.4.2.e Expectation of helpfulness of treatment

Four trials provided data on expectation (UKBEAM, Heymans et al, Moffett et al and Thomas et al). The various scales used to measure expectation in these trials are shown in Table 7.8. The majority of patients seem to have high expectation of helpfulness of treatments. The ratio of high to low expectation ranged from 23:1 for combined exercise-manipulation treatment in the UK BEAM trial to 1.2:1 for low intensity back school treatment in Heymans et al trial. In the Thomas et al trial, only two patients reported low expectation of help from acupuncture and 75 (31%) said they didn't know. The only exception was in the Moffett et al trial, where a larger number of patients had low expectation of helpfulness of class exercise treatment than those who had high expectation (123 (65.8%) vs 64 (34.2%)).

Within-arm adjusted analyses (Table 7.9) did not show the association between expectation and symptom progression to be statistically significant in any trial arm for the outcomes of interest and at any follow up point. The direction of the associations was inconsistent across trial arms.

The results of the pooled analyses (Table 7.12) did not show the association between expectation and symptom progression to be significant for RMDQ at any follow up point. The direction of the associations also varied at different follow up points.

7.4.2.f Preference for treatment

Three trials provided data on pre-randomisation patients' preference for treatments (Moffett et al trial, scale: prefer active treatment or 'indifferent',

Table 7.8 Descriptive summary of expectation of helpfulness of treatments

			Expectation, n (%)			
Trial	**Expectation scale**	**Trial treatments**	**High**	**Low**	**Missing**	**Don't know**
UKBEAM 2004	Helpfulness of each treatment: Very helpful, Helpful, Not helpful	Exercise alone	958 (71.8)	146 (10.9)	230 (17.2)	
		Manipulation alone	1037 (77.7)	66 (4.9)	231 (17.3)	
		Combined exercise and manipulation	1057 (79.2)	47 (3.5)	230 (17.2)	
		GP care	650 (48.7)	454 (34)	230 (17.2)	
Moffett et al 1999	Belief in effectiveness of exercise classes: 0-5	Exercises classes	64 (34.2)	123 (65.8)	0	
		Control (Routine GP care)	-	-	-	
Heymans et al 2006	Helpfulness of each treatment: VAS 0-10	High intensity back school	175 (55)	128 (40.2)	15 (4.8)	
		Low intensity back school	156 (52.2)	128 (42.8)	15 (4.8)	
		Usual occupational care treatment	161 (53.8)	122 (40.8)	16 (5.4)	
Thomas et al 2006	Expectation of help from acupuncture: Yes, No, Don't know	Acupuncture	162 (67)	2 (0.8)	0	75 (31)
		Usual GP care	-	-	-	

Table 7.9 Within-arm differences in the change of outcome scores (mean, 95% CI) according to expectation of helpfulness of treatment (low compared with high expectation) for the outcomes of interest. Positive results indicate larger improvement in symptoms.

Trials treatment arms	Outcome	Time, weeks	Low expectation Unadjusted	Adjusted
Heymans et al 2006				
-Usual occupational care	RMDQ	13	-0.46 (-5.20, 4.28)	-1.13 (-5.90, 3.64)
		27	0.90 (-3.77, 5.58)	-1.23 (-5.93, 3.46)
		52	-1.29 (-5.43, 2.85)	-1.82 (-5.98, 2.34)
-Low intensity back school		13	-0.09 (-4.24, 4.06)	1.52 (-2.48, 5.52)
		27	-0.99 (-5.26, 3.29)	-0.04 (-4.56, 4.47)
		52	0.41 (-3.53, 4.35)	1.45 (-2.52, 5.42)
-High intensity back school		13	1.77 (-2.62, 6.16)	2.17 (-2.38, 6.70)
		27	-0.18 (-4.20, 3.85)	0.22 (-4.03, 4.47)
		52	-2.35 (-6.43, 1.74)	-2.01 (-6.15, 2.14)
-Usual occupational care	Pain 0-100	13	-0.7 (-11.8, 10.5)	-0.2 (-11.4, 11.0)
		27	2.0 (-10.9, 15.0)	2.2 (-11.5, 15.9)
		52	-3.8 (-15.8, 8.2)	-3.4 (-16.0, 9.3)
-Low intensity back school		13	-7.0 (-20.8, 6.8)	-4.8 (-20.1, 10.5)
		27	-8.1 (-22.0, 5.7)	-8.2 (-23.2, 6.8)
		52	-10.8 (-23.2, 1.6)	-8.1 (-21.1, 4.9)
-High intensity back school		13	6.7 (-6.4, 19.8)	7.7 (-5.8, 21.1)
		27	-7.5 (-23.1, 8.0)	-9.2 (-25.2, 6.9)
		52	-7.0 (-22.1, 8.0)	-6.7 (-22.6, 9.2)
Moffett				
-Exercise programme	RMDQ	13	0.85 (-0.78, 2.47)	0.71 (-0.98, 2.40)
		27	-0.24 (-2.47, 2.0)	-0.60 (-2.88, 1.68)
		52	1.0 (-1.04, 3.04)	0.89 (-1.21, 2.98)
Thomas				
-Acupuncture	ODI	13	0.44 (-4.71, 5.59)	1.12 (-4.03, 6.27)
		52	-1.70 (-7.78, 4.38)	-1.17 (-7.23, 4.90)
UKBEAM				
-Best GP care	RMDQ	13	-0.34 (-1.51, 0.83)	-0.02 (-1.25, 1.21)
		52	-0.26 (-1.55, 1.03)	0.56 (-0.77, 1.88)
-Exercise		13	-0.28 (-1.99, 1.43)	0.10 (-1.72, 1.92)
		52	0.04 (-1.80, 1.88)	0.56 (-1.32, 2.44)
-Manipulation		13	0.08 (-2.23, 2.38)	0.09 (-2.54, 2.72)
		52	0.59 (-2.02, 3.20)	1.23 (-1.63, 4.08)
-Combined exercise & manipulation		13	-1.46 (-4.54, 1.63)	-1.55 (-4.72, 1.64)
		52	-2.70 (-5.79, 0.41)	-2.70 (-5.84, 0.45)

Johnson et al trial, scale: ‘prefer active treatment’, ‘prefer control’ or ‘indifferent’ and Thomas et al trial, scale: ‘prefer acupuncture’, ‘prefer control treatment’ or ‘do not mind’). In the Moffett et al trial118 participants (63%) preferred the trial index treatment, class exercise, and the remaining 69 (37%) were ‘indifferent’. Of those who had a preference, 53 (45%) had their preference satisfied. In the Johnson et al trial, 134 participants (57%) had a preference and 100 (43%) were ‘indifferent’. 114 participants (48.7% of all trial participants) preferred active treatment and a much smaller number (20, 8.5% of all trial participants) the control treatment. Of those who had a preference (134), 71 (53%) had their preference satisfied. In the Thomas et al trial 194 (97%) preferred acupuncture, only one (0.5%) usual care, five (2.5%) ‘did not mind’ and data were not available for 39 participants (16%).

Sufficient data were available from these trials to analyse the influence of presence of preference on symptom progression. However, because of the large difference in numbers in groups in the Thomas et al trial, the related results require caution. Within-arm adjusted analyses (Table 7.10) showed that merely having a preference for a treatment (regardless of the type of treatment and whether or not the preferred treatment was received) did not influence symptom progression in the three trials for the selected outcome measures and at any follow up point. The direction of the associations was also not consistent for the two main outcome measures and across all follow up time points.

Studying the influence on symptom progression of preference satisfaction (allocation to the preferred treatment) among participants who expressed a preference, the number of participants in the groups in the Thomas et al trial was

Table 7.10 Within-arm differences in change of outcome scores (mean, 95% CI) according to the presence of preference (preference not present compared with preference present) for the outcomes of interest. Positive results indicate larger improvement in symptoms.

Trials treatment arms	Outcome	Follow up weeks	n*	Preference not present Unadjusted	Adjusted
Johnson et al 2007	**RMDQ**				
-Physiotherapist led community-		13	53:57	-0.74 (-2.49, 1.02)	-0.72 (-2.44, 1.01)
based treatment program		27	50:55	0.38 (-1.35, 2.11)	0.41 (-1.31, 2.12)
		52	48:53	-0.59 (-2.75, 1.57)	-0.61 (-2.80, 1.59)
-Back pain educational booklet		13	43:70	0.04 (-1.74, 1.83)	-0.03 (-1.82, 1.77)
and usual GP care		27	39:59	0.23 (-1.74, 2.19)	-0.01 (-2.01, 1.99)
		52	39:55	1.46 (-0.52, 3.44)	1.18 (-0.86, 3.22)
-Physiotherapist led community-	**Pain 0-100**	13	53:57	-2.59 (-11.63, 6.46)	-1.79 (-11.00, 7.42)
based treatment program		27	50:55	-4.85 (-13.89, 4.18)	-5.18 (-14.37, 4.01)
		52	49:53	-7.77 (-18.28, 2.75)	-7.62 (-18.55, 3.31)
-Back pain educational booklet		13	43:70	0.42 (-9.22, 10.06)	-0.61 (-10.42, 9.20)
and usual GP care		27	39:59	3.86 (-7.56, 15.28)	1.80 (-9.71, 13.31)
		52	39:55	2.93 (-8.29, 14.16)	0.64 (-10.90, 12.17)
Moffett et al 1999	**RMDQ**				
-Exercise programme		6	35:50	-0.94 (-2.50, 0.62)	-0.90 (-2.50, 0.71)
		27	31:46	-0.80 (-2.96, 1.37)	-0.69 (-2.85, 1.46)
		52	34:49	-0.63 (-2.58, 1.31)	-0.64 (-2.58, 1.30)
-Usual GP care		6	31:63	-.02 (-1.68, 1.64)	0.17 (-1.55, 1.89)
		27	29:57	-0.40 (-2.48, 1.68)	0.10 (-2.0, -2.20)
		52	30:58	-0.77 (-2.67, 1.13)	-0.38 (-2.32, 1.56)
Thomas et al 2006	**ODI**				
		27	3:120	-12.37 (-29.55, 4.80)	-11.69 (-29.03, 5.65)
-Acupuncture		52	3:110	-8.12 (-27.18, 10.94)	-9.31 (-28.83, 10.21)
		104	2:92	-5.00 (-24.86, 14.84)	-5.78 (-26.43, 14.86)
-Usual care		27	2:58	-9.56 (-30.72, 11.	-9.71 (-31.54, 12.11)
		52	0:48	-	-
		104	1:40	-22.59 (-56.08, 10.89)	-21.99 (-57.37, 13.39)

*number of participants whose treatment preference was not satisfied compared with that of those whose preference was satisfied.

too small to allow analyses and therefore data from this trial were not included in this analysis. The Moffett et al trial provided data only for preference for the active treatment and not for the control treatment and therefore also could not be included in this analysis. Only the Johnson et al trial provided sufficient data on preference for both treatments of the trial. For this reason, the influence of preference satisfaction among participants who expressed a preference ('indifferent' group not included) on symptom progression was only studied in the Johnson et al trial. A sensitivity analysis was carried out to investigate whether including the 'indifferent' group in the comparison group, would influence the results of the analyses, for the Johnson et al and the Moffett et al trial (Table 7.11).

The results of adjusted within-arm analyses showed that the influence of satisfaction of preference for treatment was only statistically significant in one arm (back pain educational booklet, control arm of the Johnson et al trial), and only for RMDQ at 27 weeks. The direction of this significant association, similar to that of the majority of the other non-significant associations, was consistent suggesting that patients whose preference for treatment was not satisfied were more likely to show smaller responses to treatments.

Including data for the 'indifferent' group, there was no significant association between preference satisfaction and symptom progression in any arm of the two trials. However, apart from one trial arm at only one follow up point, the direction of associations (statistically insignificant) was the same, suggesting that those who were allocated to the treatment they did not prefer or those who didn't have a

Table 7.11 Within-arm differences in change in outcome scores (mean, 95% CI) for the outcomes of interest according to preference satisfaction (not allocated to the preferred treatment compared with being allocated to it). Comparison is presented where the 'indifferent' groups were included and not included in the variable. Positive results indicate larger improvement in symptoms.

		Follow up	'Indifferent' group not included			'Indifferent' group included		
Trials treatment arms	**Outcome**	**Weeks**	**n***	**Unadjusted**	**Adjusted**	**n***	**Unadjusted**	**Adjusted**
Johnson et al 2007	**RMDQ**	13	5:52	-1.22 (-5.64, 3.21)	-1.61 (-6.0, 2.78)	58:52	-0.87 (-2.63, 0.88)	-1.02 (-2.75, 0.71)
-Physiotherapist led community-		27	5:50	-2.40 (-6.48, 1.68)	-2.88 (-6.91, 1.12)	55:50	-0.07 (-1.81, 1.66)	-0.18 (-1.90, 1.55)
based treatment program		52	5:48	-2.98 (-7.69, 1.74)	-3.46 (-8.35, 1.43)	53:48	-1.07 (-3.23, 1.08)	-1.26 (-3.45, 0.93)
-Back pain educational booklet		13	55:15	-1.39 (-4.11, 1.32)	-1.62 (-4.38, 1.13)	58:55	-1.24 (-3.79, 1.30)	-1.07 (-3.61, 1.47)
and usual GP care		27	46:13	-4.11 (-7.18, -1.03)	<u>-4.10 (-7.26, -0.93)</u>	52:46	-3.59 (-6.33, -0.85)	-3.40 (-6.19, 0.62)
		52	45:10	-3.36 (-6.59, -0.12)	-3.21 (-6.53, 0.11)	49:45	-2.40 (-5.56, 0.77)	-2.16 (-5.38, 1.06)
-Physiotherapist led community-	**Pain 0-100**	13	5:52	-14.33 (-36.95, 8.46)	-11.55(-34.34, 11.23)	58:52	-4.73 (-13.76, 4.29)	-4.02 (-13.24, 5.20)
based treatment program		27	5:50	-13.61 (-36.12, 8.90)	-14.68 (-38.10, 8.75)	55:50	-6.77 (-15.76, 2.22)	-7.50 (-16.65, 1.66)
		52	5:48	-12.18 (-36.49,12.14)	-12.82 (-38.52, 12.87)	54:48	-9.22 (-19.70, 1.26)	-9.29 (-20.22, 1.64)
-Back pain educational booklet		13	55:15	3.94 (-11.58, 19.46)	4.32 (-11.67, 20.31)	58:55	3.75 (-10.02, 17.53)	5.00 (-8.88, 18.88)
and usual GP care		27	46:13	-16.10 (-33.84, 1.64)	-15.12 (-33.28, 3.05)	52:46	-12.70 (-29.01,3.61)	-11.26 (-27.61,5.08)
		52	45:10	-12.07 (-30.13, 6.00)	-11.82 (-30.03, 6.39)	49:45	-9.69 (-27.54, 8.17)	-11.19 (-29.27,6.88)
Moffett et al 1999	**RMDQ**	6				35:50	-0.94 (-2.50, 0.62)	-0.90 (-2.50, 0.71)
-Exercise programme		27				31:46	-0.80 (-2.96, 1.37)	-0.70 (-2.85, 1.46)
		52				34:49	-0.63 (-2.58, 1.31)	-0.64 (-2.58, 1.30)

*number of participants whose treatment preference was not satisfied compared with that of those whose preference was satisfied. Underlined indicates significant ($p<0.05$).

preference, showed smaller symptom progression compared with those whose preference was satisfied.

Pooled analyses were only possible for data on the presence of a preference. The results of these analyses (Table 7.12) showed that the associations were not statistically significant between having a preference and symptom progressions. However, the direction of these non-significant associations was similar across all follow up time points, suggesting that participants who expressed a preference for a treatment were more likely to report larger improvement in function compared with those who did not have a preference.

7.4.3 Pooled analysis

The outcome of the pooled analyses shows low heterogeneity between included treatment arms as shown in Table 7.12. As an example, a forest plot of the association of episode duration and symptom progression at 52 weeks is shown in (Figure 7).

The results of the sensitivity analyses that were conducted to assess the influence of clustering of responses within trials on the associations between the non-specific factors and responses to treatment are presented in Table 7.13. The results showed that the associations did not significantly vary by trial.

Table 7.12 Pooled mean differences (random effects model) in the change in outcome scores (mean, 95% CI) for RMDQ, according to gender, age, duration of current episode of back pain, history of back pain, expectation of helpfulness of treatment and preference for treatment.

	13weeks[b]	I^2	27weeks[c]	I^2	52weeks	I^2
Gender, *male*	0.26 (-0.31, 0.84)	31%	0.57 (-0.66, 1.81)	37%	0.48 (-0.42, 1.0)	45%
Age, *<43 years*	0.17 (-0.28, 0.61)	0%	0.45 (-0.80, 1.70)	41%	0.35 (-0.13, 0.83)	0%
Duration of LBP, *<12 weeks*	0.76(0.25, 1.27)	7%	0.27 (-1.15,1.68)	46%	1.03 (0.49, 1.57)	0%
History of LBP, *absent*	-0.05 (-0.81, 0.71)	15%	0.02 (-1.76, 1.80)	0%	-0.84 (-1.56, -0.13)	0%
Expectation of the received treatment [a], *low* (Heymans et al, Moffett et al, UK BEAM)	0.16 (-0.60, 0.92)	0%	-0.47 (-2.14, 1.20)	0%	0.28 (-0.55, 1.11)	2%
Preference for treatment, *absent* (Johnson et al, Moffett et al)	-0.40 (-1.24, 0.44)	0%	-0.01 (-0.20, 0.18)	0%	-0.12 (-1.12, 0.87)	0%

[a] Expectation of helpfulness of the received treatment dichotomised as low and high. [b] In Moffett et al trial a 6 week follow up time was used to represent a 13 week follow up. [c] Data for 27 weeks analysis was provided only in three trials, Heymans et al, Johnson et al and Moffett et al.

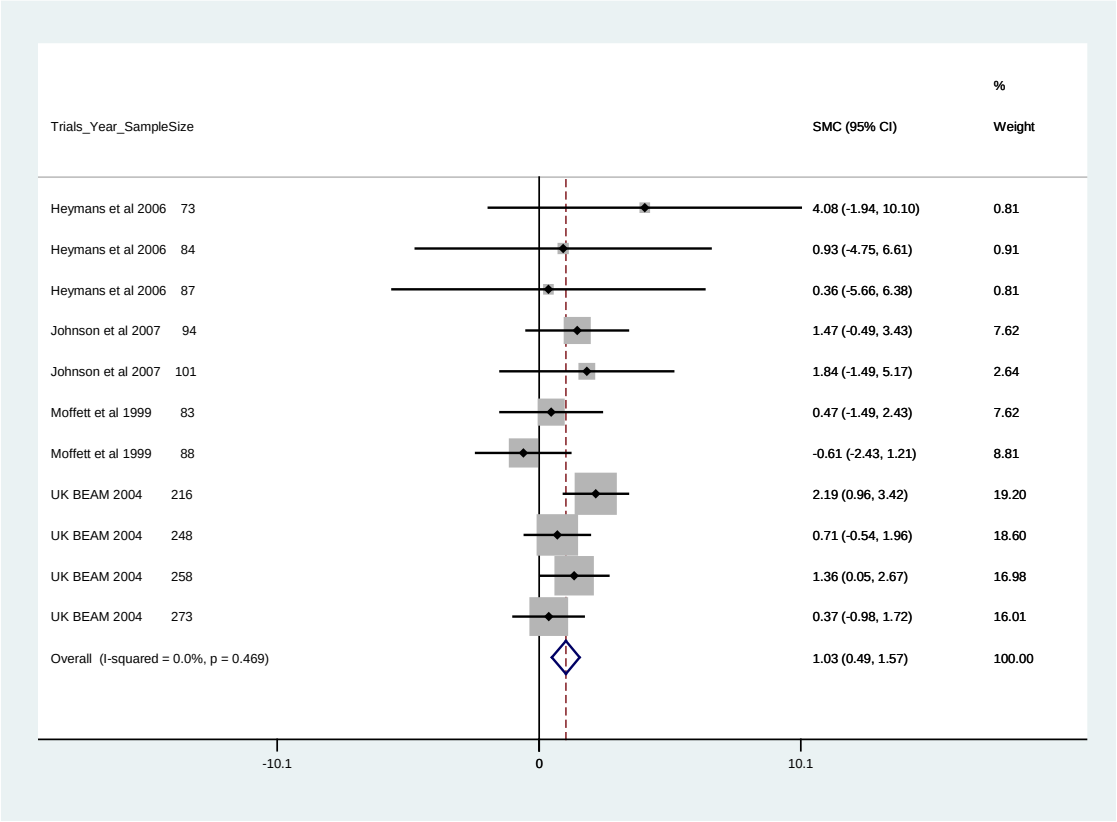


Figure 7 Mean difference in standardised mean change in RMDQ at 52 weeks between patients who had back pain episode duration of 12 weeks or less versus longer than 12 weeks. Positive indicates change associated with shorter episode duration.

Table 7.13 Pooled mean differences in change in RMDQ outcome scores according to gender, age, duration of current episode of back pain, history of back pain, outcome expectation and preference for treatment, adjusted for trial. The Moffett et al trial is the reference trial.

		13weeks	Tau^2	27weeks[a]	Tau^2	52weeks	Tau^2
Gender, male							
Unadjusted		0.28 (-0.32, 0.88)	0.00	0.57 (-0.98, 2.12)	0.97	0.25 (-0.58, 1.07)	0.46
Adjusted	Hay et al 2005	-0.54[b] (-3.50, 2.43)	0.01			0.58 (-2.11, 3.27)	0.00
	Heymans et al 2006	-1.21 (-6.31, 3.89)		-0.56 (-6.44, 5.33)	2.01	0.62 (-3.72, 4.95)	
	Johnson et al 1999	-0.12 (-2.94, 2.70)		0.54 (-4.32, 5.41)		0.13 (-2.71, 2.97)	
	UK BEAM 2004	-0.19 (-2.32, 1.94)				-1.28 (-3.35, 0.79)	
Age <43 years							
Unadjusted		0.17 (-0.32, 0.66)	0.00	0.43 (-1.17, 2.02)	0.70	0.35 (-0.18, 0.89)	0.00
Adjusted	Hay et al 2005	-0.14 (-2.58, 2.31)	0.00			0.48 (-1.94, 2.89)	0.00
	Heymans et al 2006	-0.47 (-4.03, 3.10)		-0.13 (-6.18, 5.93)	1.89	-0.31 (-3.68, 3.07)	
	Johnson et al 1999	-0.65 (-2.88, 1.59)		-0.86 (-6.29, 4.57)		0.18 (-2.42, 2.79)	
	UK BEAM 2004	-0.38 (-2.02, 1.26)				0.58 (-1.26, 2.41)	
Duration of LBP <12 weeks							
Unadjusted		0.76 (0.19, 1.33)	0.00	0.24 (-1.57, 2.05)	1.25	1.01 (0.30, 1.73)	0.19
Adjusted	Heymans et al 2006	3.48 (-1.12, 8.08)	0.00	3.92 (-1.54, 9.39)	0.19	1.86 (-2.62, 6.34)	0.13
	Johnson et al 1999	-0.17 (-2.21, 1.86)		1.98 (-1.10, 5.05)		1.68 (-1.07, 4.42)	
	UK BEAM 2004	0.69 (-0.85, 2.22)				1.29 (-0.64, 3.22)	
History of LBP, no history							
Unadjusted		-0.05 (-0.94, 0.84)	0.32	-		-0.82 (-1.66, 0.03)	0.09
Adjusted	Hay et al 2005	2.26 (-0.52, 5.04)	0.00	-		-1.23 (-4.30, 1.85)	0.00
	Heymans et al 2006	-0.41 (-4.68, 3.86)				-1.39 (-5.86, 3.08)	
	UK BEAM 2004	-0.26 (-2.53, 2.01)				-2.36 (-5.01, 0.28)	
Expectation, low							
Unadjusted		0.16 (-0.75, 1.07)	0.00	-		0.29 (-0.70, 1.27)	0.00
Adjusted	Heymans et al 2006	0.22 (-3.71, 4.16)	0.00	-		-1.62 (-6.09, 2.85)	0.00
	Johnson et al 1999	-0.81 (-3.30, 1.67)				-0.56 (-3.83, 2.71)	

[a] Data for 27 weeks analyses was provided only in three trials, Heymans et al, Johnson et al and Moffett et al trials. [b] As an example, this value means that Hay et al trial showed mean difference in response between genders 0.54 smaller than the reference trial Moffett et al, but not statistically significant.

7.5 Discussion

In this chapter, IPD analysis was used to study the influence of patient characteristics on within-arm symptom progression for back pain. Evidence was found for the influence of the duration of pain episode at inclusion in clinical trials on symptom progression. A favourable trend, although not statistically significant, was found for the influence of gender, age and presence and satisfaction of preference for treatment and symptom progression. Men, patients younger than 43 year old, patients who reported a preference and patients who were allocated to the preferred treatment seemed to be more likely to report larger responses to treatments, based on pain and functional disability. These findings will be discussed in detail.

7.5.1 Duration of back pain

Analysing pooled data from five IPD sets provided evidence for the influence of duration on symptom progression. Participants with duration of back pain of 12 weeks or less had on average up to 1 point more improvement on the RMDQ compared with those with longer duration.

It is relevant to point out that the evidence for the influence of episode duration on symptom progression within arms of each included trial was limited to a small number of arms. The advantages and disadvantages of using cut off points in analyses have been mentioned earlier in the thesis. One caveat is that the meaning of the difference between groups immediately on the two sides of the cut off point might be unclear. Also, and as has been touched on in the previous chapter, there is a growing challenge to the traditional classification of back pain

symptoms simply based on duration (Dunn et al 2008, von Korff & Dunn 2008), with the alternative proposition being that almost all back pain is a ‘chronic problem with an untidy pattern of grumbling symptoms’ (Croft et al 1998).

However, there is empirical evidence from observational studies for duration of back pain as an important prognostic factor (Henschke et al 2008, da Costa et al 2009). It is also recognised that patients who suffered from back pain for a longer duration tend to have additional associated features and characteristics that would negatively influence their symptom progressions (BenDebba et al 1997, Groth-Marnat et al 2000). Longer duration back pain has been suggested to be associated with psychological distress (Waxman et al 1998, 2000), poorer prognosis (van den Hooger et al 1997), extensive care seeking behaviour (Boutron et al 2004), more time off work and more likelihood of depression (Dunn et al 2008). For these reasons, traditional classification of back pain into acute and chronic based on duration of symptoms is still used to guide treatment recommendations in clinical guidelines, which gives utility to using it in research.

7.5.2 History of back pain

The evidence in this thesis for the influence of history of back pain on symptom progression was limited to the pooled analysis of data from a number of trials and only at 52 week follow up time.

There is evidence from the literature that the presence of a history of back pain is an important predictor of the development of a new episode (Papageorgiou et al 1996, Hestbaek et al 2003) and may also be associated with the severity of prevalent back pain. In a population survey, Hincapie et al found that a history of

occupational back pain was associated with high severity low back pain (Hincapie et al 2008). Another survey among factory workers found that previous history of back pain was associated with reduced functioning at work (Ratzon et al 2007). I am not aware, however, of any evidence for history of back pain as a predictor of symptom progression in clinical trials.

Although data on the history of episodes of low back pain were collected in five of the six included trials, this data either were not provided in trial publications (Moffett et al trial), only published as part of baseline characteristics that were used to adjust effect estimates (Hay et al, Heymans et al & Thomas et al trials) or studied among other factors as treatment effect modifiers (Underwood et al 2007). In none of the trials was the influence of history of back pain on symptom progression explored directly.

It is important to note that scales used to assess this variable in the trials varied which might have compromised the analyses. Another issue is that with the current understanding of back pain being a condition of ongoing symptoms that run a course with fluctuations, relapses and remissions in the majority of cases (Dunn et al 2006, 2008, 2010, von Korff & Dunn 2008), the distinction between ‘duration of current episode’ and ‘previous episodes of back pain’ as if they were exclusive features becomes less convincing. The validity of questions related to these features is not clear, nor is patients’ understanding and interpretation of these questions and how they answer them. All these reasons, in addition to the small number of datasets that was available for this study, might explain the failure of finding a clear evidence for the influence of history of back pain on

symptom progression, assuming it has such an influence. A clear conclusion, therefore, could not be drawn from these findings.

7.5.3 Expectations and preferences

No clear evidence was shown for the influence of expectation of helpfulness of treatment or preference for treatment on symptom progression. It is important to note the small number of trials providing data on these outcomes which reduced the power of analyses, as well as the variation in the scales used which further compromised the ability to compare or pool data from trials. There seems to be a trend that suggests those who reported a preference and who received the treatment they preferred were more likely to report larger response to the treatment. A larger volume of data would have provided more statistical power and a better opportunity to establish clear evidence on the importance of treatment preferences.

Heymans et al collected data on expectation. However, they published these data as part of baseline patients' characteristics and did not explore its influence on symptom progression. Moffett et al collected data at baseline on participants' level of belief in the effectiveness of the trial index treatment (class exercises), but similarly did not present assessment of the influence on treatment outcome in their published paper. In the UK BEAM trial, participants were asked pre-randomisation about their expectation of helpfulness for each of the trial treatments. They reported a significant association between the expectation of helpfulness of the combined treatment (exercise and manipulation) and treatment outcome at 52 weeks (Underwood 2007). They were, however, not confident of this finding commenting that it could be a chance finding.

In their published paper, Moffett et al used baseline pre-randomisation preference for treatment to adjust the reported treatment effects. They compared change in RMDQ scores at 52 weeks within each arm between participants who preferred the treatment and those who were indifferent, and found no difference between the two preference groups in the change in RMDQ scores. Johnson et al, on the other hand, used a similar approach to that adopted in this thesis where improvement in symptoms was assessed among those who received the treatment they preferred and those who didn't. They found that patients allocated to the intervention for which they had expressed a preference had clinically important reductions in pain and disability. There was no reference in the published paper of the Thomas et al trial of patient preference regarding treatment.

Similar to preference for treatment, expectation of helpfulness of treatments has been widely studied to examine its association with treatment 'effect', in back pain as well as in other medical conditions (Crow et al 1999). Evidence was not found for an association between expectation and symptom progression in this thesis. Intuitively, patients who receive their preferred treatment and who have high expectations regarding treatment would be expected to show larger response to it compared with those who do not. The evidence for this, however, is conflicting, both for expectation and preference. Some trials did not find evidence for the association between preference and expectation and treatment 'effect' (Sherman et al 2010, Leykin et al 2007) while others did (Lutz et al 1999, Kalauokalani et al 2001, Linde et al 2007, Myers et al 2008).

Preference and expectation are complex characteristics and it is plausible that the inconsistent evidence for their association with benefit of treatment might be related to the simple methods used to explore these characteristics and interpret their meaning. Various elements and factors need to be considered before simply interpreting patient's expectation that a treatment is helpful to mean that they would show a large response to it. High expectation could also reflect patient's desperation for a helpful treatment, which might represent a negative attitude with a potential negative influence on symptom progression. It has been suggested that merely having a preference indicates a level of articulation and health awareness that could be associated with higher symptom progression (Thomas et al 2006, Johnson 2007). This illustrates the complexity of investigating preference, when various explanations for conflicting findings related to various aspects of the concept of preference for treatments have been proposed.

Interactions of expectation and preference with other concepts such as general beliefs, attitude, previous use of the treatment, duration and severity of symptoms, influences of friends and family, cultural background and educational level are possible. Although preference and expectation have been used interchangeably (Johnson et al 2007) and there is some evidence to suggest a positive correlation between the two concepts (Torgeussell 1996), they convey distinct meanings. Preferring a particular treatment might not always mean that the patient has high expectation of its helpfulness, but could be because the patient wanted to try a new treatment with an open mind, prefer a specific treatment for practical reasons, or disliked other alternative treatments. Studying the correlation between preference and expectation was beyond the scope of this thesis and empirical

evidence from in-depth exploration of the meaning and reasons for preference and expectations would be valuable.

The inter-relation between patients' preference for treatment and participants blinding to treatments is also important to explore. In the two trials that explored preference for treatments, participants were not blinded to allocation to treatments and, as stated in the published trials, the main reason for exploring preference was an attempt to ameliorate the possible influence of un-blinded allocation to treatment on outcome (Moffett et al and Johnson et al trials). The majority of randomised clinical trials among back pain patients do not or cannot blind participants to trial treatments, either because it is difficult for the nature of the treatment e.g. manual treatments or the trials are pragmatic by design and participants are intentionally not 'blinded'. Trials included in this chapter were on non-pharmacological treatments, which is not unusual for primary care treatments used for non-specific low back pain (van Tulder et al 1997). The significance of this is that in trials where participants are not blinded, i.e. are able to see whether the treatment they receive is or is not their preferred, the influence of the satisfaction of their preference on symptom progressions would be most relevant. In other words, in pragmatic trials on back pain treatments in which blinding of patients is not enforced, part of the size of their symptom progression might be related to the fact that they have received their preferred treatment. This is one reason why it is worthwhile investigating the potential influence of non-specific factors in these trials in particular.

Another issue is that there is a concern that patients with a strong preference for a particular treatment might self-select or self-exclude themselves from a clinical

trial if they felt their preference would not be satisfied (Fairhurst & Dowrick 1996). This would be particularly the case in unblinded trials, in which such preferences could also lead to drop out, poor compliance and attrition (Corrigan & Salzer 2003). Although trials that explored the effect of 'un-matched' preference on drop-out rate (Leykin et al 2007) found no evidence for such an effect, more robust evidence from larger trials is required. Furthermore, to examine the impact of preference on agreeing to enrol in a trial, rate of attrition, drop out or compliance with treatment, the strength of preference, not simply the presence of preference, might be useful to assess. This is not commonly done.

7.5.4 IPD analysis

Compared with the well-developed methods for the more commonly used meta-analysis of aggregated data, there is a wide variation in the methods used for IPD meta-analysis (Simmonds et al 2005). A review conducted by Simmonds et al showed that two general methodological approaches were used in the majority of IPD reviews (Simmonds et al 2005), a one-stage and a two-stage analysis.

In the one-stage analysis, IPD from multiple trials are combined and analysed retaining original trials' identities that would be used in appropriate statistical analyses, such as a multilevel framework, to adjust for the cluster effect within each trial. In the two-stage approach data are analysed within each trial separately in the first stage and the summary statistics combined using standard summary data meta-analysis. The two-stage approach has been found to be used in the majority of IPD meta-analyses (Simmonds et al 2005). However, these two approaches were found to provide similar results (Riley et al 2010).

As the method of analysis presented in this thesis was to study responses to treatments within trial arms and not within-trial treatment effect, the 2-stage approach was adopted here. There was still the risk of the influence of cluster effect within each trial on the associations between variables and symptom progressions. To assess the potential effect of this in this thesis, the trials were included as independent variable in a meta-regression analysis as a sensitivity analysis. An alternative approach would have been to use multilevel analysis at the 2nd stage. However, given the small number of trials and small total number of arms, this was not feasible and the option that was adopted here was a less complicated and a valid alternative.

Corroborating with the results of the meta-regression analyses using AD presented in Chapter Six, duration of current episode of pain was associated with responses to treatment both on trial level using aggregated data and on patients level using individual participants data. The evidence for the association of gender with symptom progressions using AD was not supported by a clear evidence using IPD. Age was not associated with symptom progression in either analysis strengthening the evidence for this characteristic.

7.6 Limitations

Using individual data from several trials, it was possible to study the influence of patients' characteristics on symptom progressions for non-specific low back pain in more detail and a more powerful analytic approach compared with using aggregated data. Data from six trials were obtained to explore the association with

factors such as preference for treatments, expectation of helpfulness of treatment and history of back pain for which data were not available for the aggregated data analysis.

The IPD analyses were secondary analyses which were not intended in the design of the original trials. Also, although similar variables were used by a number of trials, different scales were used such as those for history of back pain and expectation. The attempts to standardise data for comparison might have led to a degree of distortion of data and affected the outcome of analyses. Extra care and caution, therefore, was taken to dichotomise data in a clear and justified way. This, however, was at the cost of reduced statistical power, and a reduced ability to compare trials or pool data from various trials. An example was data on preference for treatment which could not be pooled as data for all trial arms was not provided by all trials.

The number of datasets that could be obtained was small. This hindered conducting pooled data analysis and comparing results regarding similar variables from a large number of trials. The small number of available datasets also limited the opportunity to study the influence of patient characteristics on responses to similar types of treatments. This would have helped to understand whether factors such as expectations and preferences are generic or whether the size of their influence varies in relation to particular treatments.

A further limitation is the large number of analyses which might have led to some spurious findings of significant associations due to chance (as pointed out by Underwood et al 2007) and which need to be highlighted here.

7.7 Summary

Evidence was shown that improvement of back pain symptoms in clinical trials is influenced by duration of back pain and history of back pain episodes, regardless of the treatment received.

The lack of evidence for the association of other factors with symptom progression, such as expectation of helpfulness of treatment does not necessarily mean that no association exists. Rather it might reflect the complexity of those factors and the various ways they may influence symptom progression. It might equally reflect a lack of sufficient or appropriate data.

This study represents an attempt to utilise the value of using IPD of trials on back pain to study the influence of patients' or symptoms' characteristics on responses to treatments, which may be difficult and costly to explore in single trials. Beyond the findings and the associated limitations already pointed out, this study provides a framework and a basis for using IPD to study the influence of non-specific factors on within-arm responses to treatments.

Chapter Eight

The effect of participating in randomised controlled trials on patient outcome: comparing the course of back pain symptoms in RCTs and cohort studies.

8.1 Introduction

In the previous chapters, the course of back pain symptoms in clinical trials showed that the majority of participants show large improvement in their symptoms over time (Chapter Four). The findings in Chapter Six also showed that the variation in the size of response among trial arms was not explained by the type of treatment, according to the two classifications used. To explore sources of variation in the size of responses in these trials, the association between non-specific factors and symptom progressions in clinical trials was studied (Chapter Six and Seven). Some of those factors were related to the trials themselves while others to the patients and their symptoms.

Another non-specific factor that was suggested to influence the course of symptoms in clinical trials is the mere involvement in a trial. This is said to be related to the care and attention provided, often referred to as the ‘Hawthorne effect’, the ‘care effect’ or the ‘protocol effect’. These effects are assumed to contribute to improvement among participants in clinical trials that is in addition to the specific effects of the treatments (Braunholtz et al 2001) and have even been suggested to underline participants’ self-selection to trials (Emergency Care Research Institute website) as they seek to benefit from new treatments or treatments they perceive as effective.

If participation in a trial has a unique effect (unrelated to the treatment) it would not be expected to contribute to the variation in responses to treatments between trial arms or between trials. Rather, it would contribute to the large size of the overall symptom progressions in clinical trials in general. In other words, to

identify such an effect and be able to relate it to enrolment in RCTs, the improvement in symptoms in RCTs need to be compared with that outside the RCTs and to demonstrate a difference in the size of improvement in favour of RCTs.

The aim of this part of the study was to explore the evidence for an added improvement in back pain symptoms among RCT participants related to their participation in the trials.

8.2 Objectives

The main objective was to examine and compare the change in outcome scores over time between participants of RCTs (within-arm change of outcome scores from baseline to follow up) and those in observational cohort studies.

8.3 Methods

8.3.1 Searching and selection of studies

For randomised clinical trials, the same pool of clinical trials included in the systematic review conducted in Chapter Four was used, which were RCTs identified from the CENTRAL Cochrane register of randomised clinical trials database, accessed in April 2007.

For cohort studies, a literature search was conducted for which the detailed strategy is shown in Appendix 8. In summary, the databases of AMED, EMBASE, MEDLINE and CINAHL were searched separately using the keywords ‘low back pain’, ‘back pain’, ‘spinal pain’, ‘primary care’, ‘general practice’, ‘population’, ‘cohort’, ‘observational’, ‘prognosis’, predictor’ and ‘course’. The filters ‘human’ and ‘English language’ were used. The final search was conducted in April 2011. The results of separate searches on these databases were then combined to remove duplicates. References of relevant systematic reviews (e.g. Pengel et al 2003, Hestbaek et al 2003) and included cohort studies were checked to identify additional potentially eligible studies.

The criteria for inclusion of cohort studies were:

1. Studies: cohort or observational studies conducted in primary care (and allied services) or among the general population.
2. Participants: individuals aged 18 or over with non-specific low back pain.
3. Outcome measures: pain intensity (VAS or equivalent) and / or functional disability (e.g. RMDQ, ODI).
4. Treatments: (if any treatment used at all) primary care treatments (using similar criteria as used for RCTs).
5. Language: only studies published in English.

Excluded were studies conducted among patients with specific LBP of identifiable cause (e.g. cancer or arthritis), post-operative or post-traumatic back pain, or back pain associated with pregnancy or labour.

The result of the final combined searches was then screened and inclusion criteria applied. Full texts of potentially eligible articles were then retrieved and screened for inclusion criteria.

The literature search was conducted by the author and selection of cohort studies was checked by the supervisors (DvdW & KPJ) based on the full text of potentially eligible studies.

8.3.2 Data extraction

Data extraction was managed by two researchers, the author extracted the data which were checked by DvdW.

The extracted data included the following:

1. Study characteristics (publication year, country of study, clinical setting, sample size, drop-out rate).
2. Participants' characteristics (age; gender; duration of symptoms).
3. Outcome measures: baseline and follow up mean scores (and baseline SD) of pain intensity, RMDQ and ODI.

8.3.3 Analysis

The same three steps of analysis used in Chapter Four to study responses to treatments in RCTs were used here to study responses in RCTs and the course of symptoms in cohort studies. The steps include describing the pattern of the symptom progression in RCTs and cohort studies; assessing variation in the progression of symptoms and finally calculating pooled estimates of the change of

symptoms over time. Further to that, and for the specific objectives of this study, the change in symptoms in cohort studies was compared with that in RCT arms.

Mean outcome scores of pain and disability over time within each RCT arm and cohort study were plotted to show the pattern of symptom course. Meta-analysis, using random effects model, was performed using STATA/IC 11 software to compute pooled mean pain scores (and SD) at baseline and follow up points separately for RCTs and for cohort studies. Commonly used follow up times of 13, 27 & 52 weeks were selected for comparison. Data of other time points were considered to fall within the selected points if they were within a three-week range.

To estimate and compare the size of change in outcome scores in RCTs and cohort studies, these changes were analysed by calculating the standardized mean change (SMC) (Morris 2000). SMC was calculated for each trial arm and each cohort study by subtracting the follow-up mean score of the outcome measure from the baseline mean score and dividing by SD at baseline. Pooled SMCs and mean scores of all RCT arms and all cohort studies were calculated using standard random effects model meta-analyses methods. The concepts of random and fixed effects models have been explained in Chapter Four. To compute the 95% Confidence Intervals for response sizes, the variance of response size was calculated using the formula given in Chapter Four Analysis section. The overall change in outcome scores were investigated for heterogeneity by computing I^2 (Higgins et al 2003). Meta-regression analyses were conducted to assess the significance of the difference in the size of SMCs between RCTs and cohort studies at the selected three follow up points.

8.4 Results

8.4.1 Included studies

The search for cohort studies yielded a total of 964 citations including 31 from systematic reviews and cross references. Following exclusion of duplicates 653 records were screened for inclusion and 592 were excluded. The full texts of the remaining 61 articles were assessed for eligibility for inclusion and 33 were excluded (Figure 8.1) leaving 28 articles that fulfilled the inclusion criteria. Twenty four studies collected data on pain intensity, 13 on RMDQ and eight on ODI. However, pain intensity data useful for this analysis, i.e. mean scores for baseline and follow up points, were provided in 15 studies only, for RMDQ in six and for ODI in only two studies. The other studies provided either only baseline data or follow up data stratified by characteristics such as gender or use of healthcare service. The analysis was therefore focused on studies providing data on pain intensity measured with VAS or NRS for their entire cohort. Nine authors from studies that used pain intensity but did not provide data relevant for the analyses used in this study in their articles were approached for the mean pain intensity scores at baseline and follow up points. Four authors provided the requested scores. The final number of cohort studies considered for analyses was therefore 19. The number of RCTs included was 44 (104 treatment arms) that provided data on pain intensity as described in Chapter Four.

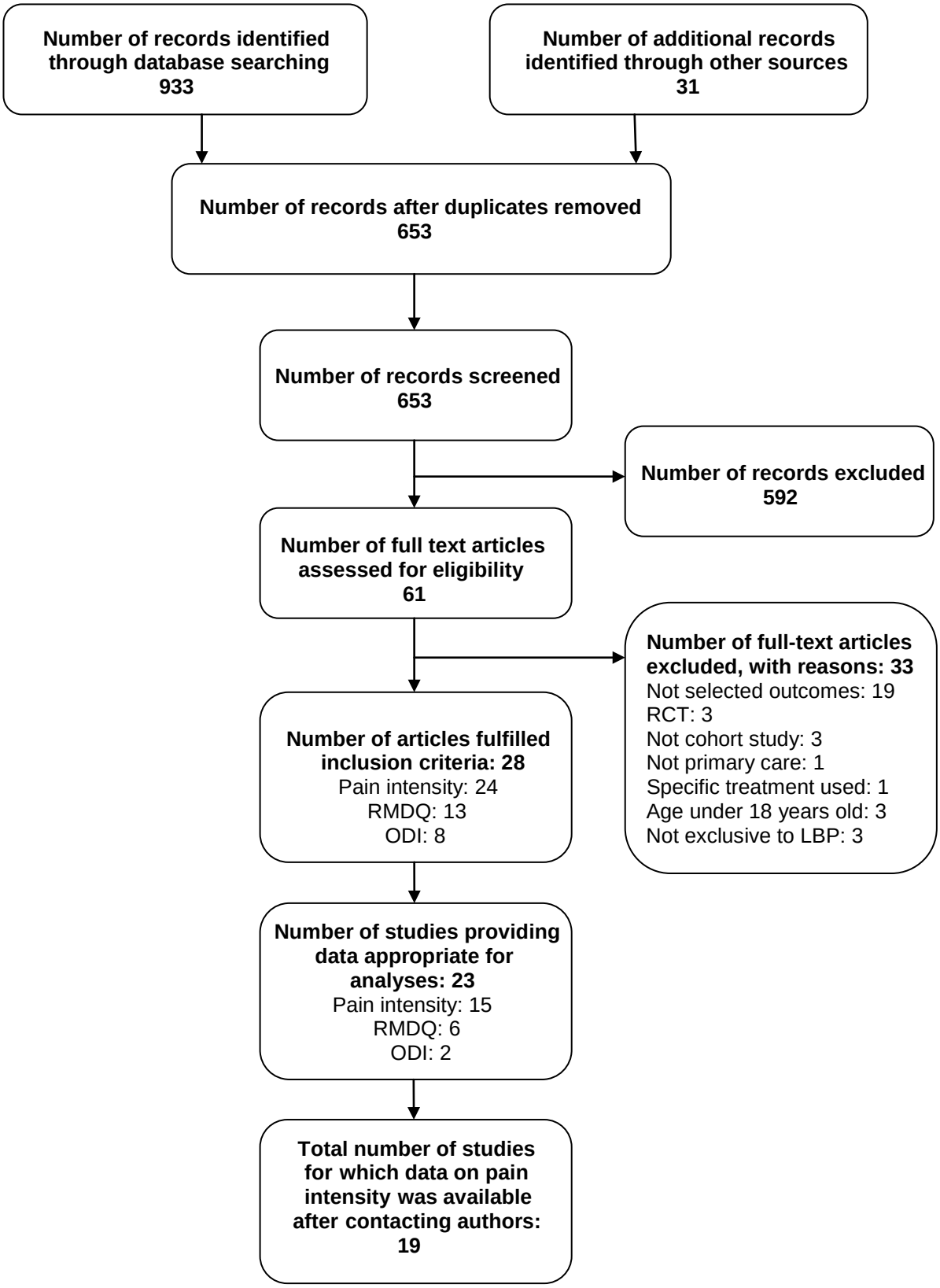


Figure 8.1 Flow chart of identification and inclusion of cohort studies in the systematic review

8.4.2 Population characteristics

Population characteristics of the included studies and trials are presented in tables 8.1 and 8.2. The trials and cohort studies were comparable for age, gender composition, mean pain duration and baseline pain intensity. Although not directly contemporaneous, the trials and cohort studies were both conducted within a similar period of time, 1993-2010.

The19 cohort studies included populations that were representative of back pain patients. They included direct cohorts from the general population as well as consulters in general practice (11 studies) and other allied primary care services such as chiropractic clinics, physiotherapy departments. All participants were described to be receiving ‘usual’ or ‘standard care’. They were conducted in 13 countries including the USA and Australia, and European countries. There was no major difference in population characteristics between cohort studies that provided data included in the analyses and those that did not (Table 8.2).

8.4.3 The pattern of pain scores over time

Pain intensity scores were provided by the included cohort studies at baseline and various follow up time points including 1, 2, 4, 6, 8, 13, 27, 36 & 52 weeks. As mentioned in the Methods section, data from cohort studies and RCTs were used in the analyses if they were within a range of three weeks around the follow up times of 13, 27 and 52 weeks. This included nine cohort studies & 58 RCT arms at 13 weeks, 10 & 59 at 27 weeks and 13 & 40 at 52 weeks. Two cohort studies had a follow up of up to 2 and 6 weeks and were not included in the analyses.

Table 8.1 Characteristics of included cohort studies

Author	Country	Sample	Drop-out rate %	Duration of pain episode	Study population	Age, mean years	Female %	Baseline pain 0-100 mean (SD)
Bakker et al 2007	The Netherlands	97	9	Acute	GP consulters	41	48	60 (15)
BaRNS	UK	206						30 (28)
Bekkering et al 2005	The Netherlands	500	4	Acute	Physiotherapy consulters	45	52	64 (21)
Carey et al 1995	USA	1628		Acute	GP and chiropractic consulters	42	52	53 (24)
Coste et al 2009	France	103	11	Acute	GP consulters	46	40	66 (18)
Chenot et al 2008	Germany	1342	1	Chronic	GP consulters	44		51 (21)
Demmelmeir et al 2010	Sweden	379		Chronic	General population	42	55	41 (16)
Grotle et al 2007	Norway	123	9	Acute	Primary care	38	55	67 (18)
Hass et al 2002	USA	2780	63	Acute	Community chiropractic clinics	43	53	51 (25)
Hoogan et al 1998	The Netherlands	443	4	Acute	GP consulters	44	55	50 (25)
Kovacs et al 2006	Spain	648	42	Acute	GP consulters	46	52	59 (22)
McGuirk et al 2001	Australia	83		Acute	GP consulters	53	57	41 (27)
Miller et al 2002	UK	211	6	Acute	GP consulters	39	60	20 (12)
Nyiendo et al 2001	USA	835	43	Chronic	Medical and chiropractic clinics			51 (24)
Perreault et al 2006	Canada	78	0	Acute	Physiotherapy departments	51		45 (23)
Sefarlis et al 1998	Sweden	60	32	Acute	GP consulters	39		51 (14)
Sharma et al 2009	USA	2872		Chronic	Medical and chiropractic clinics consulters	40	50	54 (24)
Tamcan et al 2010	Switzerland	340	13	Chronic	General population	42	50	30 (16)
van Tulder et al 1998	The Netherlands	368	11	Chronic	GP consulters	41	49	56 (29)

Table 8.2 Comparison of population characteristics of included studies

		Cohort studies		RCTs [a]
		Excluded [e]	Included	
Number		9	19	44 trials (104 arms)
Publication year		1996-2010	1994-2010	1993-2006
Sample size, Median (range)		297 (39, 974)	368 (60, 2872)	123 (28, 681)[b] 51 (12, 340)[c]
Age, mean [d] (SD)		42 (2)	43 (4.1)	43 (5.8)
Female mean %		44 (8.7)	52 (4.8)	52 (17.5)
Type of pain n (%)	**Acute**	4 (44)	8 (42)	18 (41)
	Chronic	5 (56)	11(58)	26 (59)
Baseline pain intensity, mean [c] (95% CI)			47.29 (38.62, 55.95)	46.78 (43.69, 49.88)

[a] RCTs that provided data on pain intensity outcome. [b] Sample size of RCT. [c] Sample size of arm. [d] Mean of all cohort/RCT means. [e] Studies excluded because they did not provide data appropriate for analysis in this study.

The course of pain intensity scores in the cohort studies plotted in Figure 8.2 shows a common pattern of large and rapid improvement of pain symptoms in the majority of studies within the first 13 weeks followed by very small further improvement over the period to 52 weeks. This appears to be similar to the pattern of responses to treatments in the included RCTs as shown in Chapter Four Figures 4.3a-c.

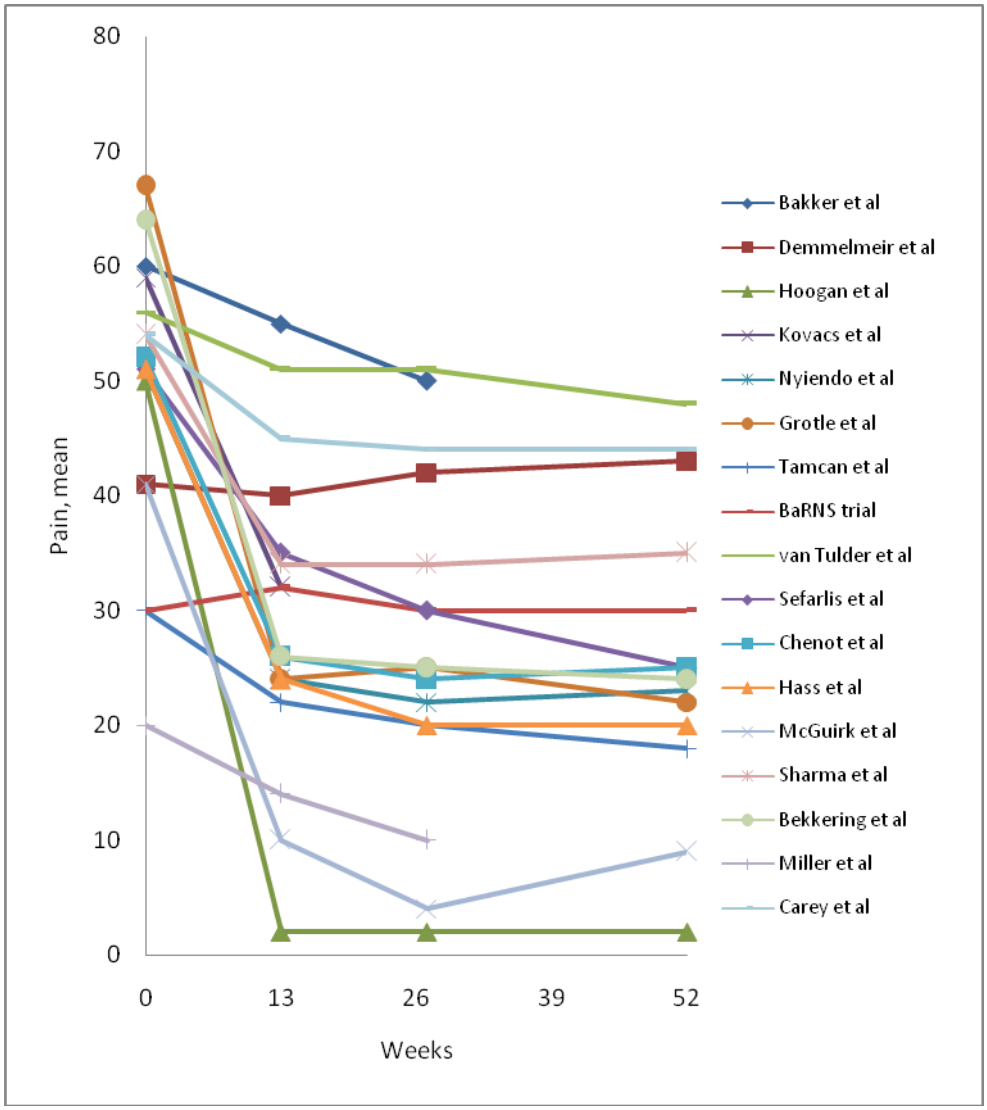


Figure 8.2 The course of pain symptoms in the included cohort studies.

8.4.4 The size of change in pain scores

Pooled mean pain intensity scores at baseline and follow up for cohort studies and RCTs are presented in Table 8.3 and Figure 8.3. The pooled estimates confirm the similar pattern of improvement in symptoms in both RCT arms and cohort studies.

Table 8.3 Pooled pain intensity scores on VAS 0-100, mean (95% CI), of included RCTs and cohort studies

	Baseline	13weeks	27weeks	52weeks
RCTs				
Pain intensity	46.78 (43.69, 49.88)	29.78 (26.63, 32.94)	25.60 (22.94, 28.26)	28.79 (25.82, 31.76)
Sample size*	7225	4820	3719	2870
Cohorts				
Pain intensity	47.29 (38.62, 55.95)	29.10 (22.72, 35.49)	25.42 (12.34, 38.51)	25.69 (20.32, 31.06)
Sample size*	13096	10965	5580	6386

*The sum of sample size of all included studies that provided data for the selected follow up times.

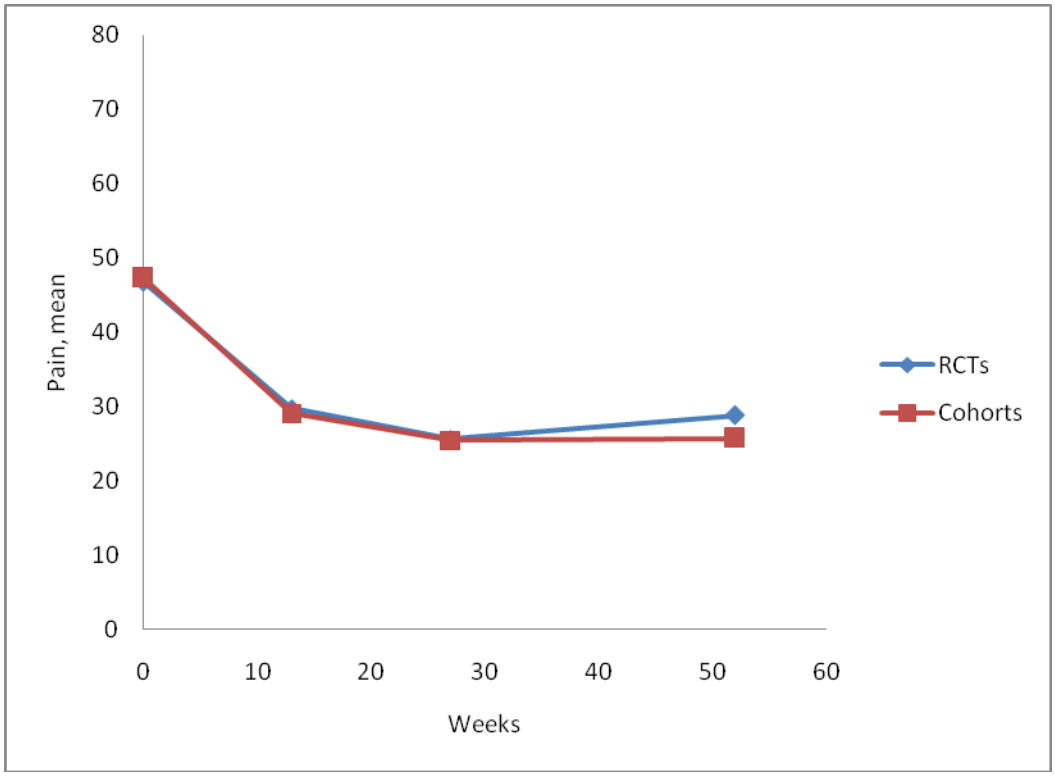


Figure 8.3 Pooled mean pain intensity scores in included cohort studies and RCTs

Change in symptoms is best represented by SMC values (Table 8.4, Figure 8.4) that take into account the distribution (standard deviation) of baseline scores.

Table 8.4 Pooled estimates of SMCs for pain intensity (baseline mean-follow up mean/baseline SD) (mean, 95% CI) of included cohorts and RCTs using random effects model meta-analysis.

	Pooled SMCs (95% CI)								
	13weeks			27weeks			52weeks		
	n		I^2	n		I^2	n		I^2
RCTs	67	0.99 (0.86, 1.11)	91	62	1.07 (0.93, 1.20)	90	55	0.91 (0.76, 1.05)	89
Cohorts	9	1.03 (0.79, 1.28)	99	10	1.27 (0.77, 1.77)	99	13	1.14 (0.81, 1.47)	99
p		0.755			0.295			0.238	

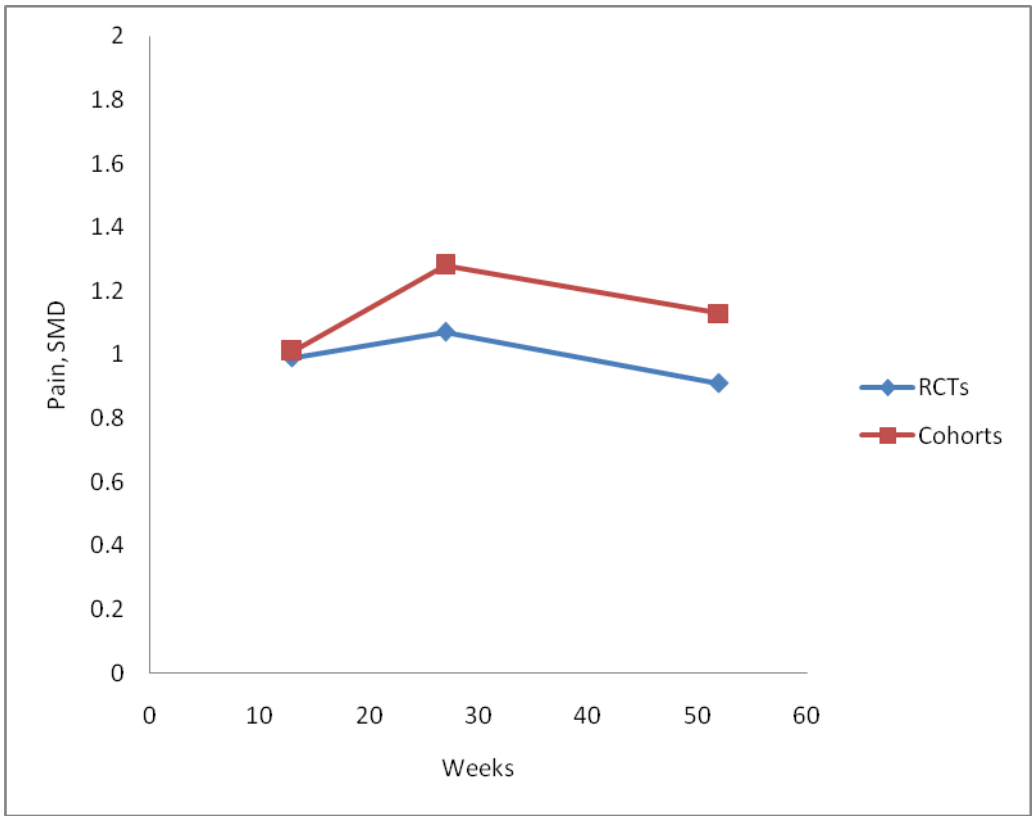


Figure 8.4 Pooled SMCs for pain in included cohort studies and RCTs

Three observations can be made from these SMC values: firstly they show large size improvement in pain symptoms in both RCT arms and cohort studies, ranging from 0.91 for RCT arms at 52 weeks (55 arms) to 1.28 for cohort studies at 27 weeks (8 studies). SMCs larger than 0.8 are considered large, 0.5 – 0.8 moderate and less than 0.5 small (Cohen 1997). Secondly, there is a large heterogeneity in the sizes of pain improvement in cohort studies (as represented by the high I^2 values), slightly larger than the heterogeneity in the responses in RCT treatment arms. Thirdly, Meta-regression analysis revealed no statistically significant difference between the SMC values of RCTs and cohort studies at any follow up point.

8.5 Discussion

The aim was to examine the course of back pain symptoms in cohort studies and compare it with that in randomized clinical trials. The supposition was a large improvement of symptoms among RCT participants that would be attributed to their participation in the RCT.

The results did not show evidence for the hypothesis that improvement in cohort studies was smaller than that in RCTs.

One explanation for the findings would be that such an effect does not exist. Another explanation could be the fact that RCTs and cohort studies are not greatly different in the attention and care provided to their participants, or at least how the care and attention are perceived by the participants. In other words, it is

plausible to assume that cohort participants would still perceive participation in cohort studies to be beneficial. This, if true, would explain the inability of this study to find smaller size of symptom improvement in cohort studies compared with RCTs.

The results also showed no significant difference in the course of symptoms and the size of change over time between RCTs and cohort studies. Similar to that in RCTs, this large improvement in symptoms over time common to the included cohort studies represents, at least in part, the effects of non-specific factors. These non-specific factors, which were discussed in Chapter Two, would explain much of the large variation that was found in the size of symptom improvement in the cohort studied.

Studying their association with the course of symptoms in these cohort studies, similar to how this was studied in RCTs in Chapter Six, would have been informative but was not feasible for the small number of studies.

To further understand the findings, a number of issues need to be discussed in more detail:

8.5.1 The benefits from participating in an RCT

It is not difficult to have a conjecture on the reasons why an RCT might have a beneficial effect on its participants. The invitation/recruitment process might have implicit positive assumptions or promises; the treatment offered might be a new treatment or the trial might be large and heavily publicized. Trial clinicians might assess patients more extensively, be better trained to deliver the treatments or

more careful and thorough in following and adhering to the trial protocol. The participants might have high expectations regarding the trial treatment or have particularly high perceived needs. However, such assumed beneficial trial effects have not been supported by solid unequivocal evidence.

A reason for the lack of solid evidence may lie in the difficulty in capturing a trial effect assuming it exists. Various approaches have been adopted to examine such an effect (Braunholtz et al 2001, Stiller 1994, Vist et al 2005) each with its own caveats and shortcomings. No single approach was considered as gold standard. The crucial issue is the choice of the comparison population that would have no exposure to any 'trial effect'. Various comparison populations have been suggested such as eligible patients who were not recruited for an RCT; eligible patients who refused to participate in an RCT; eligible patients from a geographical area other than that of the trial and eligible patients of practitioners who were not involved in the trial. Using RCTs nested in cohort studies is another example of a method that would provide an opportunity for a contemporaneous comparison of the two populations. The findings of attempts to examine the evidence for a 'trial effect' provided inconsistent results ranging from a positive trial effect (Braunholtz et al 2001) to a negative trial effect (Rochon et al 1999) through no-effect (Stiller 1994, Vist et al 2005).

One review (Stiller 1994) examined the published literature for survival rates for cancer with relation to entry into clinical trials. The review included eight studies that examined cancer survival rates in patients selected to enter RCTs and in those who were not but were included in population-based cancer registries (Ward et al 1992, Davis et al 1985, Bertelsen 1991, Karjalainen & Palva 1989,

MRC 197, Stiller et al 1989, Lennox et al 1970). It found that the survival rate of cancer patients who were selected to enter into clinical trials was higher than that of concurrent non-trial patients who were on the cancer registries, in all of the eight studies. The review acknowledged that these outcomes should be considered within the context of varying criteria for patient selection for the studies and also for the different qualities and completeness of cancer registries in different countries. The review also reported results regarding cancer patients treated at major specialist centres and teaching hospitals compared with smaller centres or non-teaching hospitals, which showed that higher survival rates among the former groups were not conclusively found in all studies. Studies varied in their inclusion criteria and in the meaning or relevance of the differences ascribed to the comparison centres and their populations.

A second review (Braunholtz et al 2001) specifically aimed to examine the evidence for an added effect of participating in a randomized clinical trial, again through examining the published literature. It included 14 articles, the majority of which were on cancer patients. This review nicely illustrates the wide variation in the choice of comparison group with potential significant impact on outcome. One of the 14 articles did not include a concurrent comparison or control group but simply compared outcome of RCT patients, who were asthmatics, with that forecast by the authors given the natural course of the disease. Two articles made indirect comparisons: between trial districts (where patients were entered into one of three trials) with control districts, and between treating physicians who were associated with Medical Research Council (MRC) trials and control physicians who were not. The remaining 11 articles compared the clinical outcomes of patients in one or more trials directly with those in at least one concurrent non-trial

control group. In three studies, the non-trial group was made up of people who had refused to participate in the trial; one study used refusers as well as eligible but non-recruited patients of recruiting clinicians; one study used all non-randomized patients of recruiting clinicians. They found that six trials reported larger statistically significant overall improvement among RCT participants compared with non-participants and three reported a favourable trend that was not statistically significant. This review was updated in 2004 (Peppercorn et al 2004), including seven of the earlier trials and adding a further 17, in an attempt to capture a larger data source and provide clearer evidence. Their conclusion was that sufficient data were still not available to confidently provide evidence for a trial effect.

Vist et al conducted a systematic review to compare the outcomes of participants in RCTs with non-participants who received the same or similar treatments (Vist et al 2005). They included five RCTs and 50 cohort studies. Participants of the RCTs were randomised to whether they had the option to participate. Two RCTs randomised patients to “n of 1 trial” compared with standard practice—that is, randomised, double blind, multiple crossover comparisons of an active drug with a placebo in a single patient. One study was in patients who had or had not been informed that they were in an RCT. One study in patients randomised to an RCT compared results with patients randomised to a patient preference trial in which they had a choice of treatment. Another study was in patients randomised to an RCT compared with those who were not invited to participate. The outcomes and the questions asked varied, one study measured spontaneously self-reported side effects in patients who had or had not been informed that they were in an RCT, another reported satisfaction among patients randomised to an RCT compared

with patients randomised to a patient preference trial in which they had a choice of treatment, another reported pain reduction among patients randomised to an RCT compared with those who were not invited to participate. None of these studies found significant differences in outcomes between patients treated in or outside RCTs. Of the total of 73 comparisons in the cohort studies between patients participating in RCTs and patients treated outside RCTs, 59 comparisons reported no significant differences in outcome; 10 reported significantly better outcomes for patients treated in RCTs, and four reported significantly worse outcomes for patients treated in RCTs.

The overwhelming majority of trials included in these reviews were among cancer patients. We are not aware of any work that examined the influence of participation in RCTs among back pain or other musculoskeletal patients. The nature of the medical condition or the symptoms, their prognosis and outcomes will arguably have an influence on any beneficial effect, or otherwise, from participating in a trial. Therefore, there might well be some unique features that enhance or diminish the trial effect among cancer patients compared with patients with other non-life threatening conditions such as back pain. Cancer patients, for instance, might have higher hopes and expectations from participating in a trial with potentially effective or new treatments that are often not available at the time. The results of the systematic reviews examining trial effect in cancer trials, however, did not seem to identify solid evidence for such a benefit. Other factors might confound the trial effect such as the type of cancer and its prognosis and the availability of effective treatment. Vist et al (2005) commented that the effects of such possible confounders were not explored in the included studies.

8.5.2 The ‘negative’ effects of RCTs

Participation in clinical trials might have a negative impact on some participants. One relevant issue is the role of participants’ expectations regarding the treatment which might influence any preference they might have for the trial treatment. Some participants might have such strong views regarding a particular treatment that this would influence their decision whether or not to participate and consent to randomisation. In the previous chapter some evidence, albeit inconclusive, was found for a greater improvement among RCT participants who received their preferred treatment. If patients with strong preferences self-exclude themselves from RCTs, included participants might have no or weak preference regarding the treatment used and this might possibly lead to low estimates of response. If patients with strong preference participated hoping to receive the preferred treatment then were disappointed with the outcome of treatment allocation, they might either drop out or have a poorer outcome (‘resentful demoralisation’ (Berglund et al 1997)). This second scenario might in fact augment the difference between the RCT two treatment arms, i.e. increase the specific treatment effect. The implication of all these scenarios depend on the strength of participants views and preferences, the prevalence of those with strong preferences and more importantly the number of those who self-excluded themselves because of their strong views. Another factor to consider is blinding patients to treatments, which, as was discussed in Chapter Five, is not commonly performed adequately in RCTs on back pain, intentionally not performed in some pragmatic trials or not possible with some manual treatments. Preferences, expectations, allocation to treatments, randomization, patient blinding and other factors are complex and often interact with each other, hence the complexity of their positive or negative impact on participation in an RCT and subsequent patient outcomes.

8.5.3 Outcomes to measure the ‘trial effect’

Another issue that is important to address is the exact nature of the ‘trial effect’ if such an effect exists and whether the appropriate or correct tools are being used to assess it and measure it. Outcome measures used were to assess pain intensity and functional disability. The effects of the extra care and attention provided in clinical trials and the satisfaction or otherwise of receiving or not receiving the preferred treatment might be poorly captured by these measures. Other more global tools such as patient satisfaction might be better suited to measure these effects. This measure, however, does not have a universally agreed structure or definition and it is not designed specifically to measure the trial effect. Survival rate was an outcome that was commonly used in the studies and reviews that assessed the trial effect in cancer studies mentioned earlier. It is not clear whether the reported differences in survival rates reflected the trial effect or treatment effects, as most of these reviews did not adjust for type of cancer treatment. The review by Vist et al. (2005) was the only review that compared studies using the same or similar treatments, and did not identify strong evidence for a trial effect.

8.6 Limitations

A large pool of back pain RCTs with various types of treatments for back pain was accessed to explore the size of change in pain symptoms in their various arms and compare this with the change in pain intensity in cohort studies. The number of cohort studies from which relevant data were available was too small to provide similarly powerful analyses. Although the overall numbers were sufficient to

compare response sizes, the small number did not allow for useful subgroup analyses. An example of a subgroup analysis that would be useful in the context of examining a trial effect is the type of treatments used in the comparison groups.

Excluding cohort studies that did not include data relevant to the analyses used in this study might have introduced potential selection bias. However, they were not different in their population characteristics from included studies. The populations of the included cohort studies seemed to be representative of non-specific back pain patients encountered in primary care.

Although the number of cohort studies included was small, their population characteristics appear similar to those of included RCTs and also to those of the excluded cohort studies. This strengthens these analyses and subsequent conclusion.

The method adopted in this study to examine the evidence for a trial effect might, arguably, not be the best approach. As was discussed earlier, various methods have been suggested with varying claimed accuracy in capturing this effect and these produced inconsistent results. However, it was decided to make use of the material available to attempt to achieve the study aims.

8.7 Summary

Progression of back pain symptoms in cohort studies follows a pattern that is similar to that in RCTs represented by large early improvement followed by a

smaller irrelevant further improvement. Evidence was not shown for an added effect of participating in RCTs compared with cohort studies. However, it could not be confidently concluded that this is evidence for the lack of such an effect. A better understanding of the meaning and nature of the effect of participating in a clinical trial, possibly through qualitative research, and careful design of methods to assess the effect and appropriate outcome measures that accurately capture it, are needed to provide robust evidence for or against it.

Chapter Nine

Discussion and conclusions

In this chapter, the main findings from the various stages of the study will be summarised. This will be followed by a discussion of these findings within the context of the broad study aims. This will avoid repeating the detailed discussion points that were addressed in previous chapters and instead will focus on the main overarching points. The limitations of the research will be outlined and general conclusions will be drawn and recommendations for future research made.

The overall aim of this thesis was to examine the association between non-specific factors and within-arm response (symptom improvement) to treatments for non-specific low back pain (NSLBP). The specific objectives were:

1. To examine within-arm overall responses to treatments in RCTs on non-specific low back pain to assess the pattern and the size of those responses and the extent of variation between them.
2. To examine the sources of variation in responses to treatments by investigating the association of non-specific factors (represented by trial level characteristics including mean age and gender of participants, duration of low back pain and aspects of trial quality) with overall responses to treatments in RCTs.
3. To examine the influence of patient characteristics (age, gender, duration of pain, previous history of back pain, expectation regarding helpfulness of treatment and preference for treatment) on responses to treatments using individual patient data from RCTs.
4. To examine whether merely participating in RCTs contributes to the size of symptom progression (the ‘trial effect’).

9.1 Summary of findings

9.1.1 The pattern of overall responses to treatments

Based on the results from 118 trials the thesis (Chapter Four) showed evidence for a common pattern in responses to treatments for NSLBP in RCTs, represented by a rapid and large improvement in symptoms within the initial 13 weeks after inclusion followed by a plateau showing little change in symptoms over the next 6 months. This pattern appeared to be common to most treatment arms and regardless of whether they were index or comparator treatments.

9.1.2 The association between trial characteristics and responses to treatments using aggregated data (AD)

The results of meta-regression analyses using AD of RCTs (Chapter Six) showed that male gender (for both outcome measures of pain intensity and RMDQ), short duration of low back pain episode (for pain and RMDQ), higher trial overall quality (for pain), better reported compliance (for pain), low drop-out rate (for RMDQ) and adequate patient blinding (for pain and RMDQ) were associated with larger within-arm responses to treatments.

9.1.3 The association between patient characteristics and responses to treatments using IPD

The outcome of pooled IPD analyses (Chapter Seven) showed that duration of back pain episode and history of back pain influenced responses to treatments in the six RCTs included in the analysis. Evidence was not found for a significant influence of age, gender, patient expectation of helpfulness of treatment or preference for treatment on symptom progressions using RMDQ. However, the

direction of the association between these characteristics and symptom progression for RMDQ, using both within-arm and pooled analyses, suggested a similar trend for larger responses to treatment in favour of patients younger than 43 years compared with older patients, men compared with women, and those who reported a preference for treatments compared with those who did not.

9.1.4 The ‘trial effect’

The thesis (Chapter Eight) did not show evidence for a ‘trial effect’. The size of symptom improvement as measured by pain intensity and disability in 104 treatment arms of 44 RCTs was similar to that in 19 cohort studies. Similar to RCTs there was a large variation in the size of the change in symptoms in cohort studies. However, the number of included cohort studies was too small to enable an assessment of non-specific factors as possible sources of that variation, as was performed with RCTs (Chapter Six).

9.2 Discussion of findings

9.2.1 Symptom progression

In this section the following issues will be discussed: the overall pattern of symptom progression in RCTs on NSLBP; the size of improvement in back pain symptoms in RCTs, the case definition (NSLBP) and the choice of follow up time points and of outcome measures.

9.2.1.a The pattern of responses to treatments

The findings in this thesis support the observation of a common pattern of within-arm responses to treatments in RCTs for primary care treatments for NSLBP. This provides empirical evidence for what has hitherto only been an anecdotal observation of a common pattern of responses in trial arms. There was also a large variation in the sizes of those responses. Evidence was not found that the type of treatment (index, active comparator, placebo, usual care, waiting list, pharmacological or non-pharmacological) explained that variation.

9.2.1.b The size of overall responses to treatments

Estimating the specific effects of primary care treatments for back pain was outside the objectives of this thesis. However, cautious comparison will be made between effect size estimates from the literature of treatments commonly used for NSLBP and the size of effects of some non-specific factors estimated in this thesis, in order to get some understanding of the magnitude of the size of the context effect associated with the use of treatments relative to their specific effects.

In a published systematic review which estimated the specific effects of treatments for NSLBP in RCTs (as the mean difference between response to active treatment and a placebo treatment), Machado et al (2009) reported that the treatment used in the largest number of trials was muscle relaxants and this had a pooled effect size of 12 points on a 0-100 pain intensity VAS (95% CI 7, 18) compared with placebo, from nine trials including 820 participants. Point estimates for effects of 16 treatments (half of the treatments included in the review) were no more than 10 points on a 0-100 scale for pain intensity. These treatments

included traction, physiotherapy, prolotherapy, exercise, anti-depressants, behavioural treatments, adenosine triphosphate, spinal manipulative therapy, NSAIDs, percutaneous thermo-coagulation intradiscal techniques, radiotherapy and magnets. The effect was not significantly larger than placebo for 15 of these treatments (NSAIDs being the exception).

In this thesis, the size of differences in overall responses to treatments according to a particular non-specific factor was up to 8.99 points (95% CI -1.68, 19.66) (0-100 VAS for pain) for compliance at 27 weeks using AD. Using IPD, the significant difference in within-arm symptom progression according to particular non-specific factors for pain intensity (0-100 VAS) was up to 20.1 points (95% CI 0.9, 39.4) for short versus long episode duration in the low intensity back school arm of the Heymans et al trial at 52 weeks. These values suggest that the size of effect of some non-specific factors is comparable to that of the specific effects of some treatments.

The observed small size of treatment effects in RCTs on NSLBP has a number of possible explanations. Clearly, the treatments may indeed have small specific effects. Alternatively, the specific effect may not be captured completely due to using inadequate outcome measuring tools or some non-specific factors may have large effects on symptoms that might overshadow the specific effect of the treatment. In some cases such non-specific factors might be considered an integral part of the treatment itself, e.g. communication with patient in acupuncture treatment.

9.2.1.c The nature of non-specific low back pain

A challenge in RCTs of NSLBP is the heterogeneity of the group of patients who are included in these trials as having 'non-specific low back pain'. 'Non-specific low back pain' (NSLBP) is a name that has been increasingly used since its introduction in the early eighties to refer to low back pain that has no identifiable cause or pathology. The clearly appealing simplicity this definition might provide, however, has a number of problems. First, it is not known what proportion of NSLBP patients included in RCTs has in fact an underlying specific pathology that simply has not been identified. Second, accepting that a large proportion of patients do actually have NSLBP with no specific pathology, this description encompasses patients who are widely heterogeneous in their characteristics such as severity of symptoms, history of back pain, co-morbidities and co-pharmacies, previous use of treatments, expectations, beliefs regarding illness and treatments and healthcare services utilization. Viewing these patients as a single homogeneous group and expecting them to show uniform reactions and responses to treatments, therefore, might be inappropriate. This variation in patients' characteristics is certainly not exclusive to back pain and is present among patients with other medical conditions. However, it is not clear whether the lack of an identifiable diagnosis for NSLBP might make the impact of variation in patients' characteristics on symptom progression more prominent compared with other conditions with specific diagnoses such as diabetes or hypertension. This might be explained by the arguably large heterogeneity of NSLBP patients regarding the back pain diagnosis compared to those with hypertension or diabetes. Also, although RCTs use baseline characteristics of participants to assess comparability between treatment arms, these descriptions are often restricted to socio-demographic and disease specific characteristics and do not

include the other potentially important characteristics mentioned above. Dividing non-specific back pain patients into more homogeneous subgroups with common characteristics might be one way to assess whether certain treatments might be effective in certain groups. This was the drive and the basis for the proposal of using classification systems in non-specific low back pain trials (Hay et al 2008, Flynn et al 2002). In fact, results from such trials seem to show a better response when treatments are matched with patients groups' characteristics compared with that among patients receiving unmatched treatments (Brennan et al 2006) or non-stratified care (Hill et al 2011).

9.2.1.d Follow up times in RCTs

The findings in this thesis show that the large improvement in symptoms in RCTs occurred early after the end of treatment. Findings from systematic reviews on treatments for NSLBP show that evidence for effectiveness of treatments was mainly in the short term rather than the long term (van Tulder et al 2006). In fact, in one review, treatment effects were estimated with the primary endpoint of the first follow up time after end of treatment (Keller et al 2007), citing the 'known' observation that patients improve most early in the course of trials as the justification for selecting this time point.

This raises the issue of the choice of follow up time points in RCTs. In this thesis, the follow up points of 13, 27 & 52 weeks were selected because they were the points most commonly selected by the included trials and because this allowed for a comparison of the pattern and size of response between trials. There were RCTs that provided data on earlier time points (from as early as one day after the end of treatment), however the number of such trials was too small to allow for a

useful comparison. Assessing responses to treatment in this early period is required in newly conducted trials to enable examination of the influence of non-specific factors on responses in this period.

9.2.1.e Selection of outcome measures

The detailed analyses in this thesis were focused on the outcome measures most commonly used in the included trials, primarily pain intensity for AD analyses and RMDQ for IPD analyses The usefulness of using uniform outcome measures in RCTs is widely recognised in that, among other benefits, it allows comparison and pooling of studies' findings (Jadad 1998, Jadad and Cepeda, 2000, Dworkin et al 2005).

It is relevant to discuss the issue of whether the outcome measures selected for analyses in this thesis were the 'right' outcomes that would be likely to capture the effects of non-specific factors. The two outcomes used (pain and disability) are subjective outcomes and as such would allow for and be expected to capture the effects of non-specific factors such as patient expectations and beliefs, and trial and practitioner factors. However, it could be argued that these outcome measures might not capture the totality of patient's experience with treatment. A combination of outcomes including patient perceived global improvement, patient satisfaction with treatment, depression scales and quality of life measures would be appropriate to capture the whole experience of receiving the treatment as well as the influence of other non-specific factors. The initial plan in the systematic review presented in Chapter Four was to collect data on all these outcomes in the included trials. However, there were insufficient data for most outcomes, which meant they could not be included in the analyses. In addition, the lack of

uniformity in the scales used in trials was a further obstacle for any meaningful interpretation or attempt at pooling the results of these other outcome measures.

Future researchers might consider using outcome measures that have been specifically developed based on the understanding of the complexity and interaction of non-specific factors with each other and with the treatments used. An example for the latter is a measure that incorporates expectation regarding helpfulness of treatments with reasons for these expectations (e.g. previous use of the treatment, desperation to try something new or recommendation by a friend or own practitioner) combined with preferences for treatment and reasons for the preferences. Such a tool has been developed (van Hartingsveld et al 2010) and could be used to explore the impact of patient expectations on responses to treatments.

9.2.2 The association between non-specific factors and symptom progression

AD and IPD were used to study the extent to which non-specific factors explain the variation in the size of responses to treatments. AD was used in meta-regression analyses to study factors at a trial level (trial setting, quality, participants' age and gender and duration of back pain episodes). Selection of these factors or characteristics was based on evidence for their importance from the literature and the availability of sufficient data for analyses in the included trials. IPD was used to study some of the same characteristics that were studied using AD but at the individual patient level (age, gender, pain duration). The IPD

analysis also addressed other characteristics that were considered to represent potentially important non-specific factors and for which available AD from RCTs was not sufficient for pooled analyses, such as history of back pain, patient expectation regarding helpfulness of treatment and preference for treatment.

9.2.2.a AD vs IPD

IPD meta-analysis is a lengthy and resource intensive exercise but has more statistical power compared with AD and so is more likely to detect a clinically relevant association with statistical significance. Also, as was pointed out in Chapter Six, using AD is associated with the risk of providing biased results when associations are detected that are not present within the individual study samples (aggregation bias).

9.2.2.b The associations of non-specific factors with symptom progression

Table 9.1 summarises the evidence for and the direction of the associations between the non-specific factors studied in this thesis and responses to treatments using AD and IPD analyses for the outcome measures for which data were available. Pain intensity was not used for the pooled IPD analyses because of the small number of trials that provided data for this outcome.

The table shows that for the three non-specific factors used in both AD and IPD analyses (age, gender and duration of NSLBP episode), consistently strong evidence for an association with symptom progression on pain and RMDQ outcomes was found only for duration of the back pain episode at most follow up times. The direction and size of difference in response was also consistent at these points.

Table 9.1 Summary of the evidence for the statistical significance and the direction of the associations between the studied factors and responses to treatments based on outcomes of pooled analyses of AD and IPD.

	AD						IPD		
	Pain			RMDQ			RMDQ		
	13w	27w	52w	13w	27w	52w	13w	27w	52w
Age, *<43 years*	●+	○+	○+	○+	○-	○-	○+	○+	○+
Gender, *male*	○+	●+	●+	○+	○+	●+	○+	○+	○+
Duration of NSLBP episode, *acute*	○+	●+	●+	●+	●+	●+	●+	○+	●+
History of NSLBP, *absent*							○-	○+	●-
Expectation of helpfulness of treatments, *low*							○+	○	○+
Preference for treatments									
Absent							○-	○-	○-
Treatment type *(ref: usual care)*									
Index	○-	○-	○-	○+	○-	○-			
Active comparator	○-	○-	○-	○-	○-	○-			
Waiting list	○+	○-	○-						
Placebo	○-		○-	○-					
Non - pharmacological	○+	○+		○+					
Setting *(ref: general practice)*									
Occupational health care	○-	○+	○-	●-	●-	●-			
Physiotherapy departments	○+	○+	○+		○-	○+			
General population	○-	○-	○-	●-	●-	●-			
Mixed setting	○-	○-	○-	●-	●-	○-			
Other	○+			○-	○-	○-			
Trial overall quality, *low*	○-	○-	●-		○+	○+			
				○+					
Individual quality criteria, *not adequate*									
Adequacy of randomisation	○+	○-	○-	○-	○-	○-			
Concealment of allocation to treatment arms			○-						
Patient blinding	○-	●-	●-	●-	●-	●-			
Care provider blinding	○-	○+	○-	○-		○-			
Co-interventions prevented	○-	○-	○-	○+	○+	○-			
Co-interventions standardised	○+	○+		○-					
Compliance	○-	●-	●-	○+	○-	○-			
Drop out	○-	○-	○+						
Measurements comparable	○+			○+	○+	○+			
Intention to treat analysis	○+	○-	○-	○-	○-	○-			

● Indicates statistically significant association. ○ Indicates statistically non-significant association. + Indicates higher response for given category of factor. – indicates lower response for given category of factor. Lack of a sign indicates lack of sufficient data for analysis.

An issue discussed in Chapter Seven is whether episode duration is a non-specific factor in the context studied in this thesis or whether it simply represents the natural history of back pain symptoms, with long baseline duration of symptoms reflecting chronic, persistent back pain. However, as the definition of non-specific factors adopted in this thesis covers factors that could influence symptom progressions (contributing to the clinical course of back pain symptoms) and be related to the patient, symptoms, practitioner, treatment or setting, it is reasonable to consider symptom duration a non-specific factor.

There was no conclusive evidence for the association between age and symptom progression, which was only significant using AD and only at 13 weeks. The direction of associations between age and outcome measures was also not consistent. This would suggest that responses to treatments in RCTs on non-specific low back pain do not systematically vary depending on age. This is in contrast to evidence available from some observational studies that suggest older age has a negative effect on recovery from back pain symptoms (van Doorn 1995, Bakker et al 2007).

For gender, the significant association with symptom progressions based on AD analyses for the RMDQ and pain intensity outcomes was not supported by evidence from IPD analyses. However, the direction of association was similar for both types of analyses, for both outcome measures and at all follow up time points. This would suggest a trend of larger symptom progression among men compared with women. Evidence from the literature for an association between gender and symptom progression is limited, although it suggests that female

gender is associated with higher reported pain severity and functional disability (Chenot et al 2008).

Aggregated data on history of back pain were not available from a sufficient number of RCTs to assess its association with symptom progression. Using IPD analysis, the evidence for the association between history of back pain and symptom progression was limited to the pooled IPD analysis on RMDQ at 52 week follow up. This suggests that patients with no history of back pain showed smaller responses to treatments on RMDQ only at 52 weeks. As was already discussed in Chapter Seven, the variations in the scales used in trials to collect data on this characteristic as well as the possible overlap between history of back pain and duration of back pain episode might explain the inability to arrive at a clear evidence for the association between history of back pain and symptom progression.

9.2.2.c The large unexplained heterogeneity in the size of responses

The non-specific factors studied in this thesis explained only a small proportion of the heterogeneity in symptom progression with the larger proportion remaining unexplained. This may be due to the lack of sufficient information from the published reports on important trial characteristics, as was the case for a number of the individual quality criteria or on other important factors that were not studied, such as practitioner skills, patient-practitioner interaction, patient beliefs, patient previous relevant experiences and concomitant use of other treatments.

9.2.2.d The effects of the non-specific factors in a broad context

One issue that has been discussed in Chapter Three and merits further discussion here is whether the influence of non-specific factors on symptom progression is generic and fixed across all types of patients, practitioners, treatments and settings or whether it is variable and related to the individual patient, the particular disease, symptoms or practitioner. It is plausible to expect that at least part of the effects of non-specific factors is related to or dependent on other factors.

An example of a non-specific factor that has been studied for interaction with other factors is patients' expectation regarding the treatment. Studies have suggested that expectation regarding the outcome of treatment can be influenced by patient beliefs (Kincheloe et al 1991, MacDonald et al 1980) and by verbal suggestions (Kirsch et al 1988, Benedetti et al 1999, Pollo et al 2001, Benedetti et al 2002). Treatment expectations could be positively influenced, for example, by explaining to the patient the mechanism of action of a particular treatment or by suggesting that the treatment is effective as supported by evidence. Negative expectations could be triggered by suggesting to the patient that it is not known whether the treatment offered is effective. This might be one explanation for the contradictory findings from trials that attempted to explore the effect of patient expectation on treatment outcome.

Related to this is the broad issue of the utility of identifying influential non-specific factors for research and clinical practice. Factors that influence patient outcomes have been structured around three broad categories in research terms: 1) predictors or prognostic factors, 2) treatment effect modifiers or moderators, and

3) treatment mediators (Hill & Fritz 2011). Prognostic factors are those factors that are said to influence patient outcome regardless of the treatment used. Examples include symptom duration or presence or severity of depression symptoms. Treatment effect modifiers are factors that are measured at baseline and are associated with response to a specific treatment. These include factors such as patient expectations regarding the treatment or educational level in the case of cognitive therapy. Prognostic factors and effect modifiers do not have to be causally associated with symptom progression, but can be used to identify high risk patients who may need further treatment (prognostic factors) or to select patients who may benefit from specific treatments (modifiers). 'Treatment mediators' is the name given to factors that influence or mediate the effect of a particular treatment on outcome and explain why treatments may or may not work in individual patients. Examples of such factors include self-efficacy, which may become a specific target for a treatment with the aim of increasing the effect of the treatment in particular individuals.

This is perhaps a simplified representation of a rather complex area where some factors could belong to more than one category (e.g. self-efficacy can be a prognostic factor, modifier and mediator when investigating the effects of cognitive behavioural treatment for NSLBP). Some factors could be considered as integral part of the treatment, as was mentioned earlier (e.g. reassurance or communication in acupuncture), yet may be seen as mediators, partly explaining the effect of the treatment.

9.3 Strengths and limitations

9.3.1 Strengths

3.3.1.a Multi-method approach

A strength of this thesis is the fact that its findings are based on adopting a multi-method approach in studying the data. This included systematic reviewing, meta-analyses and meta-regression analyses, using aggregated as well as individual patient data, from randomised clinical trials and cohort studies in addition to a semi-Delphi approach in a consensus study. Such approach allowed the use of a triangulating method in testing the evidence from one source of data using one approach to be corroborated with evidence from another source or another approach. On a personal educational and experience development level, which is relevant as this research project was undertaken in the context of a training fellowship, this provided me with the experience of using these various methods and sources of data and learning about their advantages and challenges.

9.3.1.b The number of included studies

Notwithstanding the fact that obtaining a larger number of studies would have provided more powerful analyses, the findings in this thesis was based on studying a large volume of data, including 118 randomised clinical trials for NSLBP in the aggregated data analysis.

9.3.1.c studying responses to a wide range of treatments

The approach in this thesis went beyond studying responses to a specific treatment, to studying responses to treatments as a generic entity, i.e. studying responses to a wide variety of treatments rather than to a particular treatment. I

am not aware of a study that adopted a similar approach in which the effects of non-specific factors on responses to treatments were studied regardless of the type of treatment.

9.3.2 Limitations

Some limitations have been referred to earlier in this chapter and in the various chapters of the thesis. Further to those, some general limitations will be outlined here:

9.3.2.a Selection of studies

Confining the source of RCTs to the CENTRAL database satisfied the specific objectives for this study in providing a sufficient pool of RCTs with a wide range of primary care treatments for NSLBP. The small number of trials that used placebo or sham treatments and trials that used pharmacological treatments is not unusual for back pain trials in primary care. Some eligible RCTs might not have been included because they were not included in the database. However, this is unlikely to have affected the outcome of the study in that there was no systematic exclusion of certain types of trials on certain primary care treatments or in certain primary care settings.

A number of limitations have already been mentioned with regard to the investigation of the trial effect (Chapter Eight). Related to the issue of selection, the choice of cohort studies as the comparator population might not be ideal to study the effect of participating in RCTs. A better method would be a comparison

of IPD of cohort studies with IPD of RCTs. More superior still would be an observational study with a nested RCT, designed specifically to examine the effect of participating in RCT.

9.3.2.b Studying single factors

The influences of non-specific factors on symptom progression are possibly more complex than is assumed when investigating the influence of single factors in isolation. Studying the factors in various combinations or incorporating the interaction between them might provide a clearer insight into their role in the clinical course of symptoms. In this thesis, studying multiple factors simultaneously was undertaken, but only in the context of adjusting for potential confounding. Further interaction analyses or including other combinations of factors were beyond the objectives of this thesis and might not be appropriate using the data available. Secondary analyses of published data carry the risk of analysing data from studies not designed for these secondary analyses. The IPD sets that were obtained for this thesis were limited, both in terms of numbers and also the non-specific factors studied, and would not have provided sufficient data for these analyses.

Novel approaches need to be developed that allow for an assessment of the interaction of these factors with each other and with the type of treatment (e.g. the influence of expectation regarding acupuncture might be different from that regarding joint injection).

9.3.2.c Other non-specific factors

Although a large number of non-specific factors were identified from the literature and from the semi-Delphi survey presented in Chapter Three, only a limited number of non-specific factors were actually studied. This is one of the possible reasons for the large remaining heterogeneity in responses to treatments that could not be explained by the effects of the studied factors. The limited data available for other important factors in the included studies precluded investigating their influence (as was identified by the Delphi survey, Chapter Three) such as patients' beliefs regarding illness and treatment, previous use of treatments, practitioner skills and communication between practitioner and patient. It is difficult, based on the thesis findings, therefore, to comment on the evidence for the importance of these factors. The large variation in the size of responses to treatment in RCTs that remained unexplained suggests that a large proportion of these responses could be related to factors other than the ones studied in this thesis. Research into the range of other factors and their influence is therefore needed to further understand their role.

9.4 Recommendations for future research

A number of recommendations could be drawn for future research, in addition to a number of recommendations that have already been mentioned earlier in this chapter such as those related to the outcome measures used and follow up times.

9.4.1 Overall symptom progression and studying non-specific factors

In this thesis non-specific factors and their effects were studied through secondary analyses of data from studies that were not designed to examine these factors. It is therefore important to corroborate the findings presented in this thesis or provide clearer evidence based on prospective studies particularly designed to examine these factors. Developing methods to prospectively study the effects of non-specific factors is outside the aims and objectives of this thesis. However, it is perhaps a natural extension to pass comments on the methods of trials to study these factors.

Methods have been used to assess the effects of the placebo treatments used in RCTs by subtracting the effects in no-treatment arms from those in the placebo arms. As was mentioned in Chapter Two, these methods would narrowly identify the 'placebo effect' of the placebo or sham treatments and not the effects of the wider range of non-specific factors. Issues related to the variety of methods used to study the 'trial effect' have also been addressed in Chapter Eight. To prospectively study the effects of non-specific factors on symptom progressions, methods have been proposed such as those by Quilty et al (2003) and Campbell et al (2005). These proposals included Zelen's design and various modifications of it, which require randomisation without prior consent in part of the study population. An example method, without a Zelen component, that could be suggested here would involve randomising patients to active and placebo treatment arms and adding a non-specific factor to examine, for example an 'enhanced consultation style' which would employ optimum communication consultation skills (Figure 9.1).

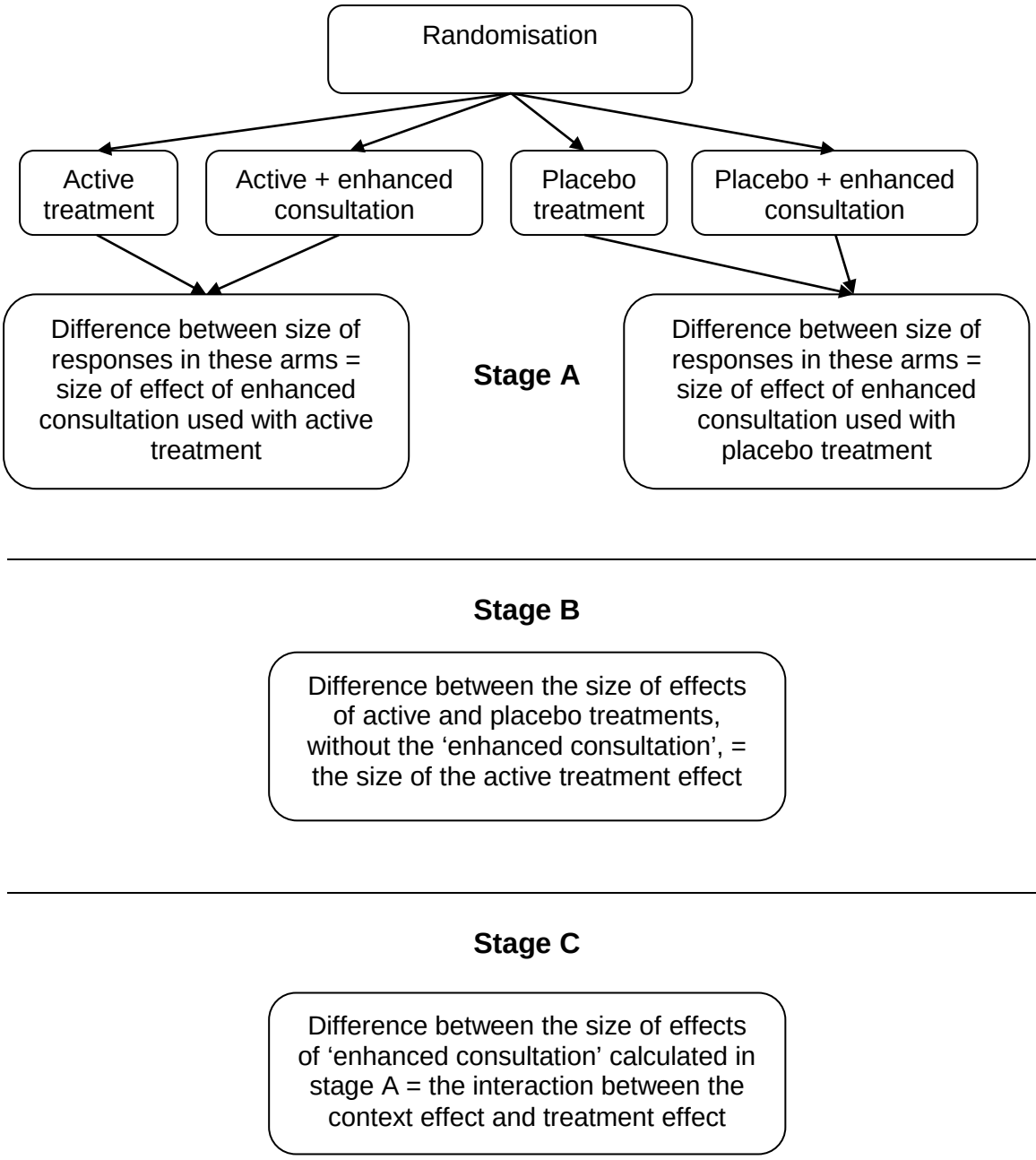


Figure 9.1 Trial design to study the effects of non-specific factors

The design would result in four arms: active treatment, active treatment with added 'enhanced consultation'; placebo treatment; placebo treatment with added 'enhanced consultation'. This should enable estimation of the effect of the enhanced consultation used in conjunction with the placebo and with the active

treatment (Stage A). The size of the effect of the active treatment is estimated as shown in Stage B. This design should also enable estimating the difference in the effects of the enhanced consultation used in conjunction with different treatments. Assuming that trial quality was adequate then any difference in the effects of enhanced consultation used in conjunction with active and with placebo (or another active treatment) (Stage C) could be attributed to the interaction between this factor and the treatment used.

Methods would need to be modified or developed to examine other non-specific factors. This illustrates the complexity of this subject area but is an important challenge for future research.

9.4.2 Duration of pain episode

Clear evidence in this thesis suggests that patients respond to treatments differently depending on the duration of the current episode. The lack of a similarly clear and consistent evidence for history of back pain might be explained by the caveats associated with the studied data, as already discussed in Chapter Seven. One recommendation that could be made, therefore, is to study (in a clinical trial) the history of back pain using a uniform scale, that would include both the presence of back pain in the past and the frequency of previous back pain (e.g. duration or number of previous episodes in a continuous variable that could be compared across trials).

The second recommendation is related to the issue that has already been discussed in Chapter Seven with relation to the degree of association between pain duration and history of back pain. It would be interesting to provide empirical

evidence for this link and to use any such evidence to design a scale to measure the two characteristics combined.

9.4.3 The quality of trials

The findings in this thesis related to the quality of RCTs on back pain raise issues that are relevant for future research. The quality of reporting of clinical trials was found to be commonly inadequate. Repeated calls have been made to improve reporting since the formal issuing of the CONSORT statement in 1996 with no clear evidence on the impact of these calls. This thesis adds further evidence for the negative impact of inadequate reporting on the ability to provide clear evidence from published trials.

The finding that the quality of a large number of back pain trials was poor might arguably be explained by the 'suitability' of quality assessment tools for different types of trials (e.g. efficacy versus pragmatic trials) and for trials using different types of treatments (e.g. pharmacological versus manual treatments). This issue, which was discussed in Chapter Five, has been recognised and attempts to modify quality assessment tools to suite a range of trials and treatments have been made. However, the trials included in this thesis were on a wide range of treatments and it might not have been possible to compare the quality across included trials had different quality tools been used.

It is pertinent to mention here that authors have historically been restricted in their publications by the size of their reports represented by the text word counts prescribed by printed journals. This might be one explanation for the less than complete reporting. The advent of the 'on-line' publications, whether through

exclusively on-line journals or complementary on-line versions of the printed journals, might present a solution for this problem and it would be interesting to establish the evidence for it in due course.

9.5 Final summary

This thesis presented evidence for the large overall improvement of back pain symptoms in clinical trials and cohort studies on back pain. The pattern of improvement was common to all types of treatment, index, active comparator, usual care, waiting list or placebo and pharmacological or non-pharmacological. The quality of studied trials was moderate with a widespread inadequate reporting of information on important aspects of methodology which did not clearly improve after the publication of the CONSORT statement in 1996.

The course of back pain symptoms in clinical trials, or the overall change in outcome scores over time, represents the effects of the specific treatment used as well as the effects of all other non-specific factors related to the patient, the symptoms, the practitioner and the setting. The approach in this thesis was to study the context effect in the context of using treatment and therefore the focus was on the overall (within-arm) symptom progressions.

The association between non-specific factors and responses to treatment was studied using published aggregated data as well as individual patient data of RCTs. Synthesised evidence was shown for the association of back pain episode duration and history of back pain with symptom progressions. Trend, but not clear

evidence, was found for the association of gender and presence of preference for treatment with symptom progression. Evidence was not found for a larger improvement in back pain symptoms in RCTs compared with cohort studies attributed to what is described as the ‘trial effect’.

The non-specific factors studied were selected based on the available data. The role of other factors, such as previous use of treatment, patient beliefs regarding illness, practitioner skills and patient-practitioner interaction, could not be assessed for the lack of relevant or sufficient data. This could explain the large variation in the size of responses to treatments that remained unexplained by the factors studied.

Obtaining a large pool of individual patient data from clinical trials on back pain would allow studying the effects of non-specific factors on within-arm responses to treatments replicating the approach adopted in this thesis. To study factors such as communication skills, newly designed prospective studies, such as the design described above, needs to be considered. Such design would also allow exploration of the possibility of the interaction between the effects of non-specific factors and the particular treatment used.

Beyond the specific findings this thesis provides, it also provides a framework for a methodological approach to study the effects of non-specific factors in clinical trials.

References

Andersson GB (1999). Epidemiological features of chronic low back pain. *Lancet*; 354: 581-5.

Balk EM, Bonis PAL, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al (2002). Correlation of quality measures with estimates of treatment effect in meta-analyses of randomised controlled trials. *JAMA*; 287: 2973-82.

Bausella RB, Leeb WL, Soekenb KL, Lic TF, Berman BM (2004). Larger effect sizes were associated with higher quality ratings in complementary and alternative medicine randomized controlled trials. *J Clin Epi*; 57: 438-446.

Beecher HK (1995). The powerful placebo. *JAMA*;159:1602-6.

Begg C, Cho M, Eastwood S, Harton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*; 276:637-9.

BenDebba M, Torgerson WS, Long DM (1997). Personality traits, pain duration and severity, functional impairment, and psychological distress in patients with persistent low back pain. *Pain*;72:115-125.

BenDebba M*, Warren S. Torgerson, Donlin M. Long (2000). A validated, practical classification procedure for many persistent low back pain patients. *Pain*; 87: 89-97.

Benedetti F, Arduino C, Amanzio M (1999). Somatotopic activation of opioid systems by target-directed expectations of analgesia. *J Neurosci*;19:3639–3648.

Benedetti F (2002). How the doctor's words affect the patient's brain. *Eval Health Prof*; 25:369–386.

Be´rard A, Bravo G (1998). Combining studies using effect sizes and quality scores: application to bone loss in postmenopausal women. *J Clin Epi*; 51, 801–807.

Berglund G, Bolund C, Gustafsson UL, Sjoden PO (1997). Is the wish to participate in a cancer rehabilitation program an indicator of the need? Comparisons of participants and non-participants in a randomized study. *Psychooncology*; 6:35-46.

Bergman S, Herrstrom P, Hogstrom K et al (2001). Chronic musculoskeletal pain, prevalence rates, and socio-demographic associations in a Swedish population study. *J Rheumatol*; 28:1369-1377.

Bertelsen K (1991). Protocol allocation and exclusion in two Danish randomised trials in ovarian cancer. *Br J Cancer*; 64:1172-1176.

Beurskens AJ, de Vet HC, Koke AJ, Regtop W, van der Heijden GJ, Knipschild PG (1995). Efficacy of traction for non-specific low back pain: a randomised clinical trial. *Lancet*; 346: 1596-1600.

Black N (1996). Why we need observational studies to evaluate the effectiveness of health care. *BMJ*; 312: 1215-8.

Borkovec TD (1985). Placebo: redefining the unknown. In: White L, Tursky B, Schwartz GE, editors. *Placebo: theory, research and mechanisms*. New York: Guilford Press; 59-66.

Boutron I, Tubach F, Giraudeau B, Ravaud P (2004). Blinding was judged more difficult to achieve and maintain in nonpharmacological than pharmacological trials. *J Clin Epi*; 57: 543-50.

Boutron I, Tubach F, Giraudeau B, Ravaud P (2003). Methodological differences in clinical trials evaluating non-pharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA*; 290:1062-70.

Branthwaite A, Cooper P (1981). Analgesic effects of branding in treatment of headaches. *BMJ* (Clin Res Ed); 282:1576-1578.

Braunholtz DA, Edwards SJL, Lilford RJ (2001). Are randomized clinical trials good for us (in the short term)? Evidence for a “trial effect.” *J Clin Epi*; 54:217-24.

Brennan GP, Fritz JM, Hunter SJ, et al (2006). Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. *Spine*; 31:623-31.

Brody H, Waters DB (1980). Diagnosis is treatment. *J Fam Pract*;10:445-9.

Brody H (1985). Placebo effect: an examination of Grunbaum’s definition. In: White L, Tursky B, Schwartz GE, editors. *Placebo: theory, research and mechanisms*. New York: Guilford Press; 37-58.

Burton AK, Tilllotson KM, Main CJ, Hollis S (1995). Psychosocial predictors of outcome in acute and subchronic low back trouble. *Spine*; 20:722-8.

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. *J Health Serv Res Policy*;10:220-5.

Carr JL, Klaber-Moffett JA, Howarth E, Richmond SJ, Torgerson DJ, Jackson DA, Metcalfe CJ (2005). A randomised trial comparing a group exercise programme for back pain patients with individual physiotherapy in a severely deprived area. *Disabil Rehabil*; 27:929-37.

Chalmers I (1989). Evaluating the effects of care during pregnancy and childbirth: in I. Chalmers, M. Enkin, M. Keirse (Eds), *Effective care in pregnancy and childbirth*. Oxford university press, London;1-37.

Charles C, Gafni A, Whelan T(1997). Shared decision- making in the medical encounter: what does it mean? (or it takes at least two to tango). *Soc Sci & Med*; 44: 681-692.

Chenot JF, Becker A, Leonhardt C, Keller S, Donner-Banzhoff N, Hildebrandt J, Basler HD, Baum E, Kochen MM, Pfingsten M (2008). Sex Differences in Presentation, Course, and Management of Low Back Pain in Primary Care. *Clin J Pain*; 24: 578-584.

Chown M, Whittamore L, Rush M, Allan S, Stott D, Archer M (2008). A prospective study of patients with chronic back pain randomised to group exercise, physiotherapy or osteopathy. *Physiotherapy*; 21-28.

The Cochrane Collaboration www.cochrane.org (accessed August 2007).

Cohen J (1997). *Statistical power analysis for the behavioural sciences*. Academic. New York.

Colloca L, Lopiano L, Lanotte M, Benedetti F (2004). Overt versus covert treatment for pain, anxiety, and Parkinson's disease. *Lancet Neurol*; 3: 679-84.

CONSORT statement http://www.consort-statement.org/ (Accessed April 2008).

Cook DJ, Sackett DL, Spitzer WO (1995). Methodologic Guidelines for systematic reviews of randomized controlled trials in health care from the postdam consultation on meta-analysis. *J Clin Epi*; 48: 167-171.

Corrigan PW & Salzer MS (2003). The conflict between random assignment and treatment preference: Implications for internal validity*. Evaluation and Program Planning*; 26, 109-121.

Coste J, Spira A, Ducimetiere P, Paolaggi JB (1991). Clinical and psychological diversity of non-specific low-back pain. A new approach towards the classification of clinical subgroups. *J Clin Epi*; 441233-45.

Coulter A (1997). Partnership with patients: the pros and cons of shared decision making. *J Health Serv Res & Pol*; 2, 112-121.

Croft PR, MacFarlane GJ, Papageorgiou AC, Thomas E, Silman AJ (1998). Outcome of low back pain in general practice: a prospective study. *BMJ*; 316:1356-9.

Crow R, Gage H, Hampson S, Hart J, Kimber A,Thomas H (1999). The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review. *Health Technol Assess*; 3(3).

da 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*; 339:c3829.

Davis S, Wright PW, Schulman SF, Hill LD, Pinkham R-D, Johnson LP, Jones TW, Kellogg HB, Radke HM, Sikkema WW, Jolly PC, Hammar SP (1985). Participants in prospective, randomized clinical trials for resected non-small cell lung cancer have improved survival compared with nonparticipants in such trials. *Cancer*; 56:1710-1718.

de Saintonge M, Herxheimer A (1994). Harnessing the placebo effects in healthcare. *Lancet*; 344: 995–98.

De Villiers M, De Villiers P, Kent AP (2005). The Delphi technique in health sciences education research. *Med Teach;27:639-643.*

DerSimonian R, Charette LJ, McPeek B, Mosteller F (1982). Reporting on methods in clinical trials. *N Engl J Med*; 306;1332-7.

Der Simonion R, Laird N (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*; 7: 177-188.

Deyo RA (1998). Low-back pain. *Sci Am*; 49-50.

Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J (2001).Influence of context effects on health outcomes: a systematic review. *Lancet*; 357: 757-62.

Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S (2002). Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof*; 25: 38-64.

Dorn SD, Kaptchuk TJ, Park JB, Nguyen LT, Canenguez K, Nam BH, Woods KB, Conboy LA, Stason WB, Lembo AJ (2007). A meta-analysis of the placebo response in complementary and alternative medicine trials of irritable bowel syndrome. *Neurogastroenterol Motil*; 19:630-7.

Dunn KM, Croft PR, Main CJ, Von Korff M (2008). A prognostic approach to defining chronic pain: Replication in a UK primary care low back pain population. *Pain*; 135: 48-54.

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

Dunn KM (2010). Extending conceptual frameworks: life course epidemiology for the study of back pain. *BMC Musculoskeletal Disorders*; 11:23.

Dworkin R, Turk D, Farrar J et al (2005). Core outcome measures for chronic pain clinical trials: IMPACT recommendations. *Pain*;113:9-19.

Egger M, Smith G, Altman DG (2009). Systematic reviews in health care. *Meta-analysis in context*. BMJ Books, London Part II, pp141.

Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G (1997). Language bias in randomised controlled trials published in English and German. *Lancet*; 350:326-329.

Elliott AM, Smith BH, Penny KI, Smith WC, Chambers WA (1999). The epidemiology of chronic pain in the community. *Lancet*; 354: 1248-1252.

The Emergency Care Research Institute website. Patients' reasons for participation in clinical trials and effect of trial participation on patient outcomes. www.ecri.org/patient_information/patient_reference_guide/evidence.pdf (Accessed Nov 2010).

Evers S, Marziniak M, Frese A, Gralow I (2009). Placebo efficacy in childhood and adolescence migraine: an analysis of double-blind and placebo-controlled studies. *Cephalalgia*; 29:436–44.

Fairbank JCT, Davies JB, Mbaot JC, O'Brien JT (1980). The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy*; 66:271-3.

Fairhurst K, Dowrick C (1996). Problems with recruitment in a randomized controlled trial of counselling in general practice: Causes and implications. *J Health Serv Res Policy*; 1:77-80.

Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP (2011). Validity of four pain intensity rating scales. *Pain*; 152:2399-404.

Finnerup NB, Sindrup SH, Jensen TS (2010). The evidence for pharmacological treatment of neuropathic pain. *Pain*; 150:573-81.

Flynn T, Fritz J, Whitman J et al (2002). A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine*; 27:2835-43.

Foster NE, Thomas E, Hill JC, Hay EM (2010). The relationship between patient and practitioner expectations and preferences and clinical outcomes in a trial of exercise and acupuncture for knee osteoarthritis. *Eur J Pain;* 14: 402-409.

Fritz JM, George S (2000). The use of classification approach to identify subgroups of patients with acute low back pain: inter-rater reliability and short term treatment outcomes. *Spine*; 25:106-114.

Frost H, Lamb SE, Doll HA, Carver PT, Stewart-Brown S (2004). Randomised controlled trial of physiotherapy compared with advice for low back pain. *BMJ*; 329;708-714.

Furlan AD, Pennick V, Bombardier C, van Tulder M, from the Editorial Board of the Cochrane Back Review Group (2009). Updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine*; 34:1929-1941.

Furukawaa TA, Barbuib C, Ciprianib A, Brambillac P, Watanabea N (2006) Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epi*; 59: 7-10.

Gallagher TJ, Hartung PJ, Gerzina H, Gregory Jr SW, Merolla D (2005). Further analysis of a doctor–patient nonverbal communication instrument. *Patient Education and Counselling*; 57: 262-271.

George SZ, Fritz JM, Childs JD, Brennan GP (2006). Sex differences in predictors of outcomes in selected physical therapy interventions for acute low back pain. *J Ortho & Sports Phys Ther*; 36:354-363.

Goodman CM (1987). The Delphi technique: a critique. *J Adv Nursing; 12: 729-734.*

Gorawara-Bhat R, Cook MA, Sachs GA (2007). Nonverbal communication in doctor–elderly patient transactions (NDEPT): Development of a tool. *Patient Education and Counselling*; 66: 223-234.

Gotzsche PC (1994). Is there logic in the placebo? *Lancet*; 344:925-6.

Gregoire G, Derderian F, Le Lorier J (1995). Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias? *J Clin Epi*; 48:159-163.

Gremeaux V, Coudeyre E, Givron P, Hérisson C, Pélissier J, Poiraudeau S, Bénaïm C (2007). Qualitative evaluation of the expectations of low back pain patients with regard to information gained through semi-directed navigation on the Internet. *Annales de réadaptation et de médecine physique*; 50:348-355.

Groth-Marnat G, Fletcher A (2000). Influence of neuroticism, catastrophizing, pain duration, and receipt of compensationon short-term response to nerve block treatment forchronic back pain. *J Behav Med*; 23:339-350.

Grünbaum A (1985). Explication and implications of the placebo concept. In: White L, Tursky B, Schwartz GE, editors. Placebo: theory, research and mechanisms. New York: Guilford Press; 9-36.

Grünbaum A (1981). The placebo concept. *Behav Res Ther*; 19:157-67.

Guyatt GH, Townsend M, Berman LB, Keller JL (1987). A comparison of Likert and visual analogue scales for measuring change in function. *J Chr Dis*; 40:1129-233.

Hardy RJ, Thompson SG (1998). Detecting and describing heterogeneity in meta-analysis. *Stat. Med*; 17:841-856.

Hasson F (2000). Research guidelines for the Delphi survey technique. *J Adv Nursing;32: 1008-1015.*

Häuser W, Bartram-Wunn E, Bartram C, Reinecke H, Tölle T (2011). Systematic review: Placebo response in drug trials of fibromyalgia syndrome and painful peripheral diabetic neuropathy—magnitude and patient-related predictors. *Pain*; 152:1709–1717.

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

Hay EM, Mullis R, Lewis M, Vohora K, Main CJ, Watson P, Dziedzic KS, Sim J, Minns Lowe C, Croft P (2005). Comparison of physical treatments versus a brief pain management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet*; 365: 2024-30.

Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, et al (2008). Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ*;337-171.

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

Heymans MW, de Vet HCW, Bongers PM, Knol DL, Koes BW, van Mechelen W (2006). The Effectiveness of High-Intensity *Versus* Low-Intensity Back Schools in an Occupational Setting. *Spine*; 31: 1075–1082.

Higgins J, Green S (2008). *Cochrane handbook for systematic reviews of interventions*. Wiley-Blackwell. Cochrane books services. Ch9, pp276.

Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003). Measuring inconsistency in meta-analyses. *BMJ*; 327:557-560.

Higgins JPT, Green S. Cochrane handbook for systematic reviews of intervention. Version 5.01. www.cochrane.org/resources/handbook/handbook.pdf (Accessed 30.07.10).

Hill CL, La Valley MP, Felson DT (2002). Discrepancy between published report and actual conduct of randomised clinical trials. *J Clin Epi*; 55: 783-786.

Hill JC, Fritz JM (2011). Psychosocial influences on low back pain, disability, and response to treatment. *Phys Ther*; 91:712–721.

Hill JC, Whitehurst DG, 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. *Lancet*; 378:1560-71.

Hincapié CA, Cassidy D, Côté P (2008). Is a history of work-related low back injury associated with prevalent low back pain and depression in the general population? *BMC Musculoskeletal Disorders*; 9:22.

Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G (1996). Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. *Med Educ*; 30:38–43.

Hoiriis KT, Pfleger B, McDuffie FC, Cotsonis G, Elsangak O, Hinson R, Verzosa GT (2004). A randomised clinical trial comparing chiropractic adjustments to muscle relaxants for subacute back pain. *J Manip Physiol Ther*; 27:388-98.

Hozo SP, Djulbegovic B ,Hozo I (2005). Estimating the mean and variance from the median, range, and the size of a sample, *BMC Medical Research Methodology*; 5:13.

Hróbjartsson A, Gøtzsche PC (2010). Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev*; 20; CD003974.

International Classification of Functioning, Disability and Health (2004). http://www3.who.int/icf/icftemplate.cfm (Accessed 11 February 2008).

Jackson JL, Chamberlin J, Kroenke K (2001). Predictors of patient satisfaction. *Soc Sci Med*; 52: 609-620.

Jackson JL, Kroenke K. The effect of unmet expectations among adults presenting with physical symptoms. *Ann Intern Med* 2001;134:889-897.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*; 17:1-12.

Jadad AR (1998). *Randomised controlled trials: a user's guide.* BMJ Books, London. http://www.bmjpg.com/rct/contents.html (Accessed 20 January 2010).

Jadad AR, Cepeda MS (2000). Ten challenges at the Intersection of clinical research, evidence-based medicine, and pain relief. *Ann Emerg Med* ;36:247–52.

Jellema P, Van der Windt DAWM, Van der Horst HE, Twisk JWR, Stalman WAB, Bouter LM (2005). Should treatment of (sub) acute low back pain be aimed at psychosocial prognostic factors? Results of a cluster-randomized clinical trial in general practice. *BMJ*; 331:84-88.

Johnson RE, Jones GT, Wiles NJ, Chaddock C, Potter RG, Roberts C, Symmons DPM, Watson PJ, Torgerson DJ, Macfarlane GJ (2007). Active exercise, education, and cognitive behavioural therapy for persistent disabling low back pain. A randomized controlled trial. *Spine;* 32: 1578-1585.

Juni P, Holenstein F, Sterne J, Bartlett C, Egger M (2002). Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epi*; 31:115-123.

Juni P, Witschi A, Bolch R, Egger M (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*; 282;1054-60.

Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA (2001). Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. *Spine*; 26:1418-24.

Kaplan SH, Greenfield S, Ware JE Jr (1989). Assessing the effects of physician–patient interactions on the outcomes of chronic disease. *Medical Care* ; 27(Suppl 3): S110–S127.

Karjalainen K, Malmivaara A, Mutanen P, Roine R, Hurri H, Pohjolainen T (2004). Mini-intervention for subacute low back pain: two-year follow-up and modifiers of effectiveness. *Spine*; 29:1069-76.

Karjalainen S, Palva I (1989). Do treatment protocols improve end results? A study of survival of patients with multiple myeloma in Finland. *BMJ*;299:1069-1072

Kassirer JP, Campion EW (1994). Peer review: crude and understudied, but indispensable. *JAMA*; 272:96-97.

Katon W, Kleinman A. Doctor-patient negotiation and other social science strategies in patient care. In: Eisenberg, L., Kleinman, A (1981). *The relevance of social science for medicine*. Reidel Publishing Company, London; 253-279.

Katz J, Pennella-Vaughan J, Hetzel RD, Kanazi GE, Dworkin RH (2005). A randomized, placebo-controlled trial of bupropion sustained release in chronic low back pain. *J Pain*; 6: 656-661.

Keller A, Hayden J, Bombardier C, van Tulder M (1991). Effect sizes of nonsurgical treatments of non-specific low-back pain. *Eur Spine J*;16:1776-88.

Kincheloe JE, Wallace MLJ, Mattison GD, Seib K (1991). Psychophysical measurement on pain perception after administration of a topical anaesthetic. *Quintessence Int*; 22:311-15.

Kirsch I, Weixel LJ (1988). Double-blind versus deceptive administration of a placebo. *Behav Neurosci*;102:319-323.

Kjaergard LL, Villemsen J, Gluud C (2001). Reported methodologic quality and discrepancies between large and small randomized trials in metaanalyses. *Ann Intern Med*;135:982-9.

Kleijnen J, de Craen AJM, van Everdingen JA, Krol L (1994). Placebo effect in double-blind clinical trials: a review of interactions with medications. *Lancet*; 344: 1347-49.

Koes BW, Malmivaara A, van Tulder MW (2005). Trend in methodological quality of randomized clinical trials in low back pain. *Clin Rheum*;19: 529-539.

Krismer M, van Tulder M (2007). Low back pain (non-specific). *Clin Rheum*; 21:77-91.

Kunz R, Oxman AD (1998). The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*; 317:1185 - 90.

Lambeek LC, van Mechelen W, Knol DL, Loisel P, professor, Anema JR (2010). Randomised controlled trial of integrated care to reduce disability from chronic low back pain in working and private life. *BMJ*; 340:c1035.

Lambert PC, Sutton AJ, Abrams KR, Jones DR (2002). A comparison of summary patient-level covariates in meta regression with individual patient data meta-analysis. *J Clin Epi*; 55: 86-94.

Lau J, Loannidis PA, Schmid CH (1998). Summing up evidence: one answer is not always enough. *Lancet*; 351:123-127.

Leboeuf C, Grbnstvedt A, Borge JA, Lothe J, Magnesen E, Nilsson A, Rbsok G, Stig L-C, Larsen K (2005). The Nordic back pain subpopulation program: a 1-year prospective multicentre study of outcomes of persistent low-back pain in chiropractic patients. *J Manip Physiol Ther*; 28:90-96.

Lennox EL, Stiller CA, Morris Jones PH, Kinner Wilson LM (1979). Nephroblastoma: treatment during 1970-3 and the effect on survival of inclusion in the first MRC trial. *BMJ*; 567-569.

Leykin Y, DeRubeis RJ, Gallop R, Amsterdam JD, Shelton RC, Hollon SD. The Relation of Patients' Treatment Preferences to Outcome in a Randomized Clinical Trial. *Beh Ther*; 38:209-217.

Licciardone JC, Brimhall AK, King LN (2005). Osteopathic manipulative treatment for low back pain: a systematic review and meta-analysis of randomized controlled trials. *Musculoskeletal Disorders*;6:43.

Licciardone JC, Stoll ST, Fulda KG, Russo DP, Siu J, Winn W, Swift Jr J (2003). Osteopathic Manipulative Treatment for Chronic Low Back Pain. A Randomized Controlled Trial. *Spine*; 28:1355-1362.

Linde K, Witt CM, Streng A, Weidenhammer W, Wagenpfeil S, Brinkhaus B, Willich SN, Melchart D (2007). The impact of patient expectations on outcomes in four randomized controlled trials of acupuncture in patients with chronic pain*. Pain; 128:264-271.*

Lutz GK, Butzlaff ME, Atlas SJ, Keller RB, Singer DE, Deyo RA (1999). The relation between expectations and outcomes in surgery for sciatica. *J Gen Intern Med* ; 14:740-744.

MacDonald AJ, Peden NR, Hayton R, Mallinson CN, Roberts D, Wormsley KG (1980). Symptom relief and the placebo effect in the trial of an anti-peptic drug. *Gut*; 21:323-6.

Macedo A, Baños JE, Farré M (2008). Placebo response in the prophylaxis of migraine: a meta-analysis. *Eur J Pain*; 12:68-75.

Maniadakis N, Gray A (2000). The economic burden of back pain in the UK. *Pain*; 84: 95-103.

Medical Research Council (MRC) (1971). Duration of survival of children with acute leukaemia. *BMJ*; 4: 7-9.

Miller FG, Rosenstein DL (2006). The nature and power of the placebo effect. *J Clin Epid*; 59:331-335.

Mills E, Wu P, Gagnier J, Heels-Ansdell D, Montori VM (2005). An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *J Clin Epid*; 58:662-667.

Moerman DE, Jonas WB (2002). Deconstructing the Placebo Effect and Finding the Meaning Response. *Ann Intern Med*; 136:471-476.

Moffett JK, Richardson G & Sheldon TA (1995). *Back pain: its management and costs to society*. Discussion Paper 129. York: Centre for Health Economics, University of York.

Moffett JK, Torgerson D, BellSyer S, Jackson D, LlewlynPhillips H, Farrin A, Barber J (1999). Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. *BMJ*; 319:279-83.

Moher D, Deborah J Cook, Susan Eastwood, Ingram Olkin, Drummond Rennie, Donna F Stroup, for the QUOROM Group (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet*; 354:1896-900.

Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S (1995). Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*; 16:62-73.

Moher D, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A (2000). What contributions do languages other than English make on the results of meta-analysis? *J Clin Epid*; 53:964-972.

Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP (1998). Does the quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*; 352:609-13.

Moher D, Schulz KF, Altman DG, for the CONSORT Group (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*; 357:1191-94.

Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH (2002). In the dark The reporting of blinding status in randomized controlled trials. *J Clin Epi*; 787-790.

Morris SB (2000). Distribution of the standardised mean change effect size for meta-analysis on repeated measures. *Bri J of Math and Stat Psych*; 53;17-29.

Morton V, Torgerson DJ (2005). Regression to the means: treatment effect without the intervention. *J Eval Clin Pract*; 11:59-65.

Mostellier F, Gilbert JP, MCPeek B (1980). Reporting standards and research strategies for controlled trials –agenda for the editor. *Control Clin Trials*; 1:37-58.

Myers SS, Phillips RS, Davis RB, Cherkin DC, Legedza A, Kaptchuk TJ, Hrbek A, Buring JE, Post D, Connelly MT, Eisenberg DM (2008). Patient expectations as predictors of outcome in patients with acute low back pain. *J Gen Intern Med*; 23:148-153.

Natvig B, Bruusgaard D & Eriksen W (2001). Localized low back pain and low back pain as part of widespread musculoskeletal pain: two different disorders? A cross-sectional population study. *J Rehab Med*; 33:21-25.

NICE guidelines for the early management of persistent non-specific low back pain (2009). www.nice.org.uk

Ong LML, de Haes JCJM, Hoos AM, Lammes FB (1995). Doctor–patient communication: a review of the literature. *Soc Sci Med*; 40:903-18.

Ostelo RW, Stomp-van den Berg SG, Vlaeyenz JW, Woltersy PM, de Vet HC (2003). Health care provider's attitudes and beliefs towards chronic low back pain: the development of a questionnaire. *Man Ther*; 8(4): 214–222.

Oxman AD, Guyatt GH (1991). Validation of an index of the quality of review articles. *J Clin Epi*; 44:1271-1278.

Papageourgiou AC, Croft PR, Ferry S, Jayson MIV, Silman AJ (1995). Estimating the prevalence of low back pain in the general population. Evidence from the South Manchester Back Pain Survey. *Spine*; 20: 1889-94.

Papageorgiou AC, Croft PR, Thomas E, Ferry S, Malcolm IV Jayson MIV, Silman AJ (1996). Influence of previous pain experience on the episode incidence of low back pain: results from the South Manchester Back Pain Study. *Pain*; 66:181-185.

Papageourgiou AC, Silman AJ, Macfarlane GJ (2002). Chronic widespread pain in the population: a seven year follow-up study. *Ann Rheum Dis*; 61:1071-4.

Paterson C, Dieppe P (2005). Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ*; 330:1202-5.

Pengel HM, Maher CG, Refshauge K (2002). Systematic review of conservative interventions for subacute low back pain. *Clin Rehab*; 16:811-820

Pengel L, Herbert R, Maher C, Refshauge K (2003). Acute low back pain: systematic review of its prognosis. *BMJ*; 327:323-8.

Peppercorn JM, Weeks JC, Cook EFC, Joffe S (2004). Comparison of outcomes in cancer patients treated within and outside clinical trials: conceptual framework and structured review. *Lancet*; 363:263-70.

Petrie KJ, Frampton T, Large RG, et al (2005). What do patients expect from their first visit to a pain clinic? *Clin J Pain*; 21(4):297-301.

Pitz M, Cheang M, Bernstein CN (2005). Defining the predictors of the placebo response in irritable bowel syndrome. *Clin Gastro Hepatol*; 3:237-47.

Pollo A, Amanzio M, Arslanian A, Casadio C, Maggi G, Benedetti F (2001). Response expectancies in placebo analgesia and their clinical relevance. *Pain*; 93:77-84.

Price D (2001). Assessing placebo effects without placebo groups: an untapped possibility? *Pain*; 201-203

Quebec Task Force on Spinal Disorders (1987). Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. *Spine*; 12(Suppl 1):1-59.

Ratzon NZ, Jarus T, Cats A (2007). The relationship between work function and low back pain history in occupationally active individuals. *Dis Reh*; 29:791-6.

Reiss S (1980). Pavlovian conditioning and human fear: an expectancy model. *Behav Ther*; 11:380-396.

Rescorla RA (1988). Pavlovian conditioning: it's not what you think it is. *Am Psychol*; 43:151-160.

Riley RD,Lambert PC, Abo-Zaid G (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*; 340:c221.

Rochon PA, Binns MA, Litner JA, Litner GM, Fischbach MS, Eisenberg D, Kaptchuk TJ, Stason WB, Chalmers TC (1999). Are randomized control trial outcomes influenced by the inclusion of a placebo group? a systematic review of non-steroidal anti-inflammatory drug trials for arthritis treatment. *J Clin Epi*; 52(2):113-22.

Roland M, Morris R (1983). A study of the natural history of back pain. Development of a reliable and sensitive measure of disability in low back pain. *Spine*; 8:141-4.

Royston P, Douglas GA, Sauerbrei W (2006). Dichotomising continuous predictors in multiple regression: a bad idea. *Stat.Med*; 25:127-141.

Ruoff GE, Rosenthal N, Jordan D, Karim R, Kamin M (2003). Tramadol / Acetaminophen combination tablets for the treatment of chronic lower back pain: a multicenter, randomized, double-blind, placebo-controlled outpatient Study. *Clin Ther*; 25:1123-1141.

Sanders SH, Brena SF, Spier CJ, Beltrutti D, McConnell H, Quintero O (1992). Chronic low back pain patients around the world: cross-cultural similarities and differences. *Clin J Pain*; 8: 317-323.

Schmid CH, Stark PC, Berlin JA, Landias P, Lau J (2004). Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level factors. *J Clin. Epi*; 57: 683-697.

Schouten BC, Meeuwesen L, Harmsen HAM (2005). The impact of an intervention in intercultural communication on doctor–patient interaction in The Netherlands. *Patient Education and Counselling*; 58: 288–295.

Schouten BC, Meeuwesen L, Tromp F, Harmsen HAM (2007). Cultural diversity in patient participation: The influence of patients' characteristics and doctors' communicative behaviour. *Patient Education and Counselling*; 67: 214-223.

Schultz KF, Grimes DA, Altman DG, Hayes RJ (1996). Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ*; 312: 742-754.

Schulz KF, Chalmers I, Grimes DA, Altman DG (1994). Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynaecology journals. *JAMA*; 272:125-8.

Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995). Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effect in controlled trials. *JAMA*; 273:408-12.

Shapiro A, Morris L (1978). The placebo effect in medical and psychological therapies. In: Garfield S, Bergin A, editors. *Handbook of psychotherapy and behavioural change*. New York: Wiley; 369-410.

Shapiro AK (1960). Attitudes toward the use of placebos in treatment. *J Nerv Ment Dis*;130:200-11.

Shapiro AK (1961). Factors contributing to the placebo effect. *Am J Psychother* ;18:73-88.

Shapiro AK (1968). The placebo response. In: Howells JG, editor. *Modern perspectives in world psychiatry*. London: Oliver and Boyd; 596-619.

Shaw A, Ibrahim S, Reid F, Ussher M, Rowlands G (2009). Patients' perspectives of the doctor–patient relationship and information giving across a range of literacy levels. *Patient Education and Counselling*; 75:114-120

Sherman KJ, Cherkin DC, Ichikawa L, Avins AL, Delaney K, Barlow WE, Khalsa PS, Deyo RA, MD, MPH (2010). Treatment Expectations and Preferences as Predictors of Outcome of Acupuncture for Chronic Back Pain. *Spine*; 35:1471-1477

Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials*; 2:209-17.

Smith BH, Hopton JL, Chambers WA (1999). Chronic pain in primary care. *Family practice*; 16; 5: 475-482.

Smith SK, Dixon A, Trevena L, Nutbeam D, McCaffery KJ (2009). Exploring patient involvement in healthcare decision making across different education and functional health literacy groups. *Soc Sci & Med*; 69: 1805-1812.

Somerville S, Hay EM, Lewis M, Barber J, van der Windt D, Hill J, Sowden G (2008). Content and outcome of usual primary care for back pain: a systematic review. *BJGP*; 58: 790-797.

Stewart MA (1995). Effective physician-patient communication and health outcomes: a review. *Can Med Assoc J*; 152:1423-33.

Stiller CA, Draper GJ (1989). Treatment centre size, entry to trials, and survival in acute lymphoblastic leukaemia. *Arch Dis Child*; 64: 657-661.

Stiller CA (1994). Centralised treatment, entry to trials and survival. *Br J Cancer*; 70:352-62.

Thomas KB (1994). The placebo in general practice. *Lancet*; 344:1066-67.

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*, doi 10.1136/bmj.38878.907361.7C.

Thompson SG, Higgins JP (2002). How should meta-regression analyses be undertaken and interpreted?. *Stat Med*; 21: 1559-1573.

Thompson SG (2009). In *systematic reviews in health care. Meta- analysis in context*. BMJ Books, London, Part II, pp 157.

Thompson SG (1994). Why sources of heterogeneity in meta-analysis should be investigated. *BMJ*; 309:1351-1355.

Torgeussell IT (1996). Patient preferences in randomised trials: threat or opportunity? *J Health Services Res Policy*; 1:194-197.

Turner JA, Deyo RA, Loeser JD, Von Korff M, Fordyce WE (1994). The importance of the placebo effects in pain treatment and research *JAMA*; 271: 1609–14.

Turpin DL (2005). Editor-in-Chief CONSORT and QUOROM guidelines for reporting randomized clinical trials and systematic reviews. *Am J Orthod Dentofacial Orthop*; 128:681-6.

UK BEAM trial team (2004). United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *BMJ*; 329: 1377-80.

Underwood MR, Morton V, Farrin et al on behalf of the UK BEAM trial team (2007). Do baseline characteristics predict response to treatment for low back pain? Secondary analysis of the UK BEAM dataset. *Rheumatology*;46:1297-1302.

van den Hoogen, HJM, Koes BW, Devillé W, van Eijk JTM, Bouter LM (1997). The Prognosis of Low Back Pain in General Practice. *Spine*; 22: 1515-1521.

van Doorn TWC (1995). Low back disability among self-employed dentists, veterinarians, physicians and physical therapists in the Netherlands. *Acta Orthop Scand*; 66 (suppl 263): 1-64.

van Hartingsveld F, Ostelo RW, Cuijpers P, de Vos R, Riphagen II, de Vet HC (2010). Treatment-related and patient-related expectations of patients with musculoskeletal disorders: a systematic review of published measurement tools. *Clin J Pain*; 26(6):470-88.

van Houwelingen JC, Arends LR, Stijnen T (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*; 21:589-624.

van Tulder M, Furlan A, Bombardier C, Bouter L (2003). Updated Method Guidelines for Systematic Reviews in the Cochrane Collaboration Back Review Group. *Spine*; 28;1290-1299.

van Tulder MW, Assendelft WJJ, Koes BW (1997). Method guidelines for systematic reviews in the Cochrane collaboration review group for spinal disorders. *Spine*; 22: 2323–2330.

Van Tulder MW, Koes BW, Bombardier C (2002). Low back pain. *Best Practice and Research Clinical Rheumatology*; 16:761-75.

van Tulder MW, Koes BW, Malmivaara A (2006). Outcome of non-invasive treatment modalities on back pain: an evidence-based review. *Eur Spine J*; 15 Suppl 1: S64-81.

van Tulder MW, Marika Suttorp M, Morton S, Bouter LM, Shekelle P (2009). Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine*; 34:1685–1692.

Verbeek J, Sengers MJ, Riemens L, et al (2004). Patient expectations of treatment for back pain: a systematic review of qualitative and quantitative studies. *Spine*; 29:2309-2318.

Verhagen A, de Vet H, de Bie R, Kessels A, Boers M, Knipschild P (1998). Balneotherapy and quality assessment: interobserver reliability of Masstricht criteria list and the need for blinded quality assessment. *J Clin Epi*; 51;335-341.

Verhagen AP, de Vet HC, de Bie RA, Bores M, van den Brandt PA (2001). The art of quality assessment of RCTs included in systematic reviews. *J Clin Epi*; 54: 651-654.

Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, Knipschild PG (1998). The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epi*; 51:1235-41.

Vist GE, Hagen KB, Devereaux PJ, Bryant D, Kristoffersen DT, Oxman AD (2005). Systematic review to determine whether participation in a trial influences outcome. *BMJ*; 330: 21.

Von Korff M, Dunn KM (2008). Chronic pain reconsidered. *Pain*; 138:267-276.

von Korff M (1994). Studying the natural history of back pain. *Spine*; 19:2041S-6S.

Waddell G (2005). Subgroups within ‘non-specific’ low back pain. *J Rheumatology*; 32:395-6.

Ward LC, Fielding JWL, Dunn JA, Kelly KA (1992). The selection of cases for randomised trials: a registry survey of concurrent trial and non-trial patients. *Br J Cancer*; 66: 943-950.

Wasan AD, Kong J, Pham LD, Kaptchuk TJ, Edwards R, Gollubz RL (2010). The impact of placebo, psychopathology, and expectations on the response to acupuncture needling in patients with chronic low back pain. *J Pain*; 11: 555-63.

Waxman R, Tennant A, Helliwell P (2000). A prospective follow-up study of low back pain in the community. *Spine*; 25:2085-90.

Waxman R, Tennant A, Helliwell P (1998). Community survey of factors associated with consultation for low back pain. *BMJ*; 317:1564-7.

White L, Tursky B, Schwartz GE (1985). Placebo in perspective. In: White L, Tursky B, Schwartz GE, editors. *Placebo: theory, research and mechanisms*. New York: Guilford Press; 3-7.

Wijnhoven AH, de Vet HC, PhD, Picavet HS (2006). Prevalence of musculoskeletal disorders is systematically higher in women than in men. *Clin J Pain*; 22:717-724

Wijnhoven HA, de Vet HC, Picavet HS (2007). Sex differences in consequences of musculoskeletal pain. *Spine*; 32: 1360-1367.

World Health Organisation. The burden of musculoskeletal conditions at the start of the new millennium. i-x (2003). *World Health Organ Tech Rep Ser*; 919: 1-218.

Zenz M, Strumpf M (2007). Redefining Appropriate Treatment Expectations. *J Pain Symptom Manage*; 33:S11-S18.

Zhang W, Robertson J, Jones AC, Dieppe PA, Doherty M (2008). The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. *Ann Rheum Dis*; 67:1716-23.

Appendices

Appendix 4.2a Summary of trials (n 118) included in the systematic review presented in Chapter Four. Citations of trials are detailed in Appendix 4.2b

Author and country	Setting	Age (y)	Treatment	Age (y) mean (SD)	Female (%)	Actual duration of NSLBP (weeks) mean (SD)	Sample of trial arm at baseline
Andersson et al, USA	Members of a health organisation	20-59	Osteopathic manipulative treatment	29 (10.6)	59		83
			Standard medical care	37 (11)	56		72
Bannwarth et al, France	General practice	18-55	Adrenaline triphosphate tablets 90 mg	43 (9.9)	59	7.4 (2)	81
			Placebo tablets	41 (8.9)	58	7 (2)	80
Bendix et al, Denmark	General practice		Functional restoration (PT + OT + Psychological)	40	66		48
			Outpatient intensive physical training: Aerobics + strengthening exercises + fitness machines	43	69		51
Beurskens et al, Netherlands	Physiotherapy + General practice	18 +	Continuous motorised lumbar traction	39 (10)	44		
			Sham traction	42 (11)	43		
Birbara et al, USA	Not clear	18-75	Etoricoxib tablets 60 mg	52 (13.3)	63	629 (546)	103
			Etoricoxib tablets 90 mg	52 (12.4)	63	562 (561.6)	107
			Placebo tablets	51 (13.7)	55	556 (520)	109
Blomberg et al, USA	General practice + occupational healthcare	20-60	Osteopathic techniques: (manipulation, mobilisation, muscle stretching) & autotraction for some & steroid injection for non-responders				48
			Conventional treatment: mainly physiotherapy (exercise), activation and advice				53
Brennan et al, USA	Physical therapy/Rehab unit	18-65	Manipulation treatment	38 (10)	45	2	46
			Specific exercise treatment: instructed to do directional exercises	37 (11.1)	34	2.4	70
			Stabilization treatment: Trunk strengthening and stabilising exercises				
Bronfort et al, USA	College outpatient clinic	20-60	Spinal manipulation & trunk strengthening exercise	41 (10.5)	54	156	71
			NSAID & Trunk strengthening exercise	40 (8.9)	44	104	52
			Spinal manipulation & Stretching exercise	41 (9.3)	39	119.6	51
Burton et al, UK	General practice	17-70	The Back Book + usual care (GP & osteopathic care)		11		83
			The traditional Handy Hints & usual care (GP & osteopathic care)		12		79

Cambron et al & Gudavalli et al, USA	Chiropractic clinic + hospital clinic + General population	18 +	Chiropractic flexion distraction procedure	42 (11.4)	34		123
			Active trunk exercise program	41 (12.8)	41		112
Carr et al, UK	Physiotherapy + General practice + Hospital consultants		Individual physiotherapy	43 (11.2)	62		119
			Back to fitness group programme: progressive stretching & strengthening exercises & CBT	42 (10.6)	59		118
Cherkin et al, USA	General practice	20-64	Physical therapy: McKenzie approach	42 (10.7)	47		133
			Chiropractic manipulation	40 (9.4)	53		122
			Educational booklet	40 (11.2)	42		66
Childs et al, USA	Air Force	18-60	Manipulation (High velocity thrust spinal manipulation) + exercise	33 (11.2)	43	3	70
			Exercise	35 (10.6)	41	4.3	61
Chiradejnant et al, Australia	Physiotherapy		Therapist (Physiotherapist) selected mobilisation	47 (16.4)		26.3 (77)	70
			Randomly selected mobilisation	54 (16.4)		13 (40)	70
Chok et al, Singapore	Physiotherapy + Orthopaedic clinics + A/E	20-55	Physical therapy (endurance exercise at the PT department) + back hot pack	38 (9.7)	20	4 (2)	38
			Back hot pack (Home)	34 (8.1)	29	4 (2)	28
Cleland et al, USA	General practice	18-60	Lumbar spine mobilisation & exercise	40 (12.2)	71	19 (12.5)	14
			Lumbar spine mobilisation & exercise & slump stretching	39 (11.3)	69	15 (8)	16
Coats et al, USA, Canada	Not clear	18 +	Valdecoxib tablets 40 mg	48.6 (13.3)	55	603 (556.4)	148
			Placebo tablets	48.7 (12.6)	59	567 (520)	145
Constant et al, France	General practice	Unlimited	Spa therapy & usual GP care				63
			Waiting list group & usual GP care				63
Cramer et al, USA	Chiropractic clinic		Chiropractic treatment				17
			Detuned USS & cold pack + soft tissue massage				19
Damush et al 2003 (2 papers), USA	General practice	18 +	Self management program & usual care	45	72		76
			Usual care (not clearly defined)	46	75		87
Defrin et al, Israel	Occupational healthcare		Shoe insert	43 (11.8)		634 (478.4)	22
			Usual care (not clearly defined)	48 (6.9)		910 (670.8)	11
Delitto et al, USA	Physiotherapy		Extension-mobilisation, matched and specifically directed (targeted)	37 (12)	50	1 (0.9)	14
			Flexion exercises, generalised, non-specific	27 (10)	33	2 (0.9)	10
Descarreaux et al, Canada	General population		Specific exercise program (increased muscular force and extensibility of trunk and hip muscles.)	33.1	33		10

			Common back pain exercise	35	33		10
Dreiser et al, France	General practice	18-60	Diclofenac - K 12.5, 2 tablets	41 (10.9)	52		124
			Ibuprofen tablets 200 mg	41 (11.6)	48		122
			Placebo tablets, 2 tablets	41 (11.3)	53		126
Faas A, Netherlands	General practice	16-65	Usual GP care	36	41		155
			Usual GP care & Placebo USS by physiotherapist	38	42		162
			Usual GP care & Exercise instructions + advice	36	47		156
Frost et al, UK	Physiotherapy	18 +	Routine physiotherapy & advice book	42 (14.9)	58		144
			Advice from physiotherapist & advice book	40 (13)	47		142
Geisser et al, USA	University spinal programme	18-65	Manual therapy & Specific exercise (self corrections, stretching, strengthening)	39 (12.8)	67	284 (493)	26
			Sham Manual therapy & Specific exercise	39 (9.4)	56	370 (447.8)	25
			Manual Therapy & Non-specific exercise	37 (14.4)	80	370 (476.1)	24
			Sham Manual Therapy & non-specific exercise	46 (9.5)	61	284 (305.1)	25
Gemmell et al, USA	Chiropractic clinic	18-65	Activator adjustment chiropractic	54 (9.5)			14
			Meric adjustment chiropractic	52 (10.3)			16
George et al, USA	Physical therapy	18-55	Standard care physical therapy	37 (10.1)	53	4 (2.3)	32
			Fear-avoidance based physical therapy	40 (10)	62	4 (2.3)	34
Giles et al, Australia	Hospital outpatients + General practice + self referral	17 +	Spinal manipulation	39 (18.5)	49	431.6	35
			Acupuncture	38 (18.8)	44	234	34
			Medications (Celebrex, Vioxx or paracetamol) 200-400, 12.5-25, 4g	39 (8.9)	42	332.8	40
Glomsrod et al, Norway	Physicians clinics and General population	18-50	Active back school (Lectures and back exercises)	41 (6.1)	65		37
			Usual medical care	39 (6.6)	57		35
Goldby et al, UK	General practice + hospital physicians	18-65	Spinal stabilisation & Attending the back school	43 (10.7)	68		84
			Manual therapy & Attending the back school	41 (11.7)	69.9		89
			Education (Booklet: Back in action) & Attending the back school	42 (13)	67.5		40
Grunnesjo et al, Sweden	Sick-listed population + General practice		Stay active (Ortho surgeons + physiotherapists)	41 (8.5)	35	4 (3.8)	72
			Stay active & manual therapy (GPs + physiotherapists)	42 (8.9)	51	4 (3.8)	88
Guillemin et al, France	General practice		Spa therapy	59 (2.3)	65		52
			Usual GP care	58 (2.3)	56		52
Haas M, USA	General population	60 +	Chronic disease self management program (CDSMP) (workshops)	79 (7.5)	82		60

			Waiting list control	76 (7.5)	88		49
Hansen et al, Denmark	Airline	21-64	Intensive dynamic back muscle exercise	40.5	32	25.7	60
			Conventional physiotherapy (CP)	38.4	32	28.6	59
			Semi-hot pack & intermittent traction	41.9	32	28.6	61
Hay et al, UK	General practice	18-64	A brief programme of pain management (general fitness and exercise at clinic and home, explanation about pain mechanisms, distress, encouragement of positive coping strategies, overcoming fear of “hurt=harm”, and implementation of a graded return to usual activities)	40 (12.0)	50		201
			Physiotherapy including manual therapy techniques	41 (11.6)	55		201
Heymans et al, Netherlands	Occupational healthcare	18-65	Usual Dutch occupational physician care	41 (9.6)	17	35	103
			Low intensity back school	41 (10.2)	22	35	98
			High intensity back school	40 (9.5)	23	35	98
Hawk et al, USA	General population	18 +	Flexion-distraction chiropractic manipulation (FDT) & trigger point	51(14.2)	52	208	54
			Sham manipulation & effleurage	53 (15.2)	63	364	57
Helmhout et al, Netherlands	Army		High intensity treatment group: high intensity progressive resistance training of isolated lumbar extensor muscle groups	41 (10)	0		41
			Low intensity treatment group: low intensity progressive resistance training of isolated lumbar extensor muscle groups	40 (9)	0		40
Hemmila et al, Finland	Not clear		Physiotherapy				34
			Bone setting (is a form of manipulative therapy still practiced by uneducated Finnish folk healers)				44
			Exercise				35
Hseih et al, USA	General population	18 +	Joint manipulation & myofascial therapy	48 (13.7)	33	12 (7.2)	52
			Joint manipulation	47 (14)	33	12 (7.2)	48
			Myofascial therapy	49 (14.8)	33	12 (6.8)	51
			Back school	48 (13.7)	40	11 (6.6)	48
Hoiriis et al, USA	General population	21-59	Chiropractic adjustments & placebo medicine	42 (9.7)	50	3 (1.3)	34
			Muscle relaxants & sham adjustments (mimic chiropractic adjustments with respect to dialogue, visit length, and physical contact.	41 (10.1)	53	4 (1.5)	36
			placebo medicine & sham adjustments (mimic chiropractic adjustments with respect to dialogue, visit length, and physical contact.	43 (9.8)	53	4 (1.4)	40
Hurley et al, UK	Physiotherapy + General practice + self referral	18-65	Manipulation therapy (Passively move intervertebral joint within or beyond its range)	40 (11.6)	57	7.5 (3.1)	80
			Interferential therapy (Electrical stimulation)	40 (12.1)	62	7.6 (3)	80
			Manipulation & interferential therapy	41 (11.3)	60	8.3 (2.8)	80
Hurley et al, UK	Physiotherapy + General practice + self referral	18-65	Interferential therapy painful area + back book	35	62	5	18

			Interferential therapy spinal nerve + back book	35	59	7	22
			Back book only	30	45	4	22
Hurwitz et al, USA	Managed care facility	18+	Chiropractic care only	52 (16.5)	49		169
			Chiropractic care & physical modalities (Heat/cold, USS)	54 (16.8)	58		172
			Medical care (excluding physical treatment) only	49 (16.5)	47		170
			Medical care & physical modalities (Heat/cold, USS)	49 (16.7)	54		170
Hurwitz et al, USA	Network of healthcare	18 +	Chiropractic care only	52 (16.5)	49		340
			Chiropractic care & physical modalities (Heat/cold, USS)	53 (16.8)	58		340
Jellema et al, Netherlands	General practice	18-65	Minimal intervention strategy (Assessing psychosocial risks, providing information on back pain and treatments & advice on self care)	43 (11.1)	48	1.7	143
			Usual GP care	42 (12)	47	1.7	171
Kaapa H, Finland	Occupational healthcare	22-75	Multidisciplinary rehabilitation: guided, group programme. : CBT, relaxation, back school education & physical therapy	46 (7.9)	100	72 (112.5)	59
			Individual physiotherapy	47 (7)	100	63 (103.5)	61
Kankaanpaa et al, Finland	Occupational healthcare		Active rehabilitation: guided exercises in a dept + behavioural support	40 (8.5)	34		30
			Passive treatment: which they considered as minor to the active arm, e.g. massage and thermal treatment	39 (7.3)	33		24
Karjalainen et al 2003 & 2004, Finlands	General practice	25-60	Mini-intervention (Specific back exercises, reduce patient concerns & encourage physical activity)	44	59		56
			Mini-intervention & worksite visit	44	57		51
			Usual GP care	43	60		57
Katz et al, 2003& 2004, USA	General practice + hospital patients	18-75	Rofecoxib 25mg tablets	53 (13.2)	63	634 (624)	233
			Rofecoxib 50 mg tablets	53 (12.9)	62	608 (613.6)	229
			Placebo tablets	45 (12.7)	63	666 (698)	228
Katz et al, USA	Hospital pain clinic + general population	18 +	Bupropion SR tablets 150 mg	50 (10)	57		26
			Placebo tablets	51 (11.4)	39		28
Kerr et al, UK	General practice		Acupuncture	43 (11.5)	50	86 (84.9)	30
			Placebo TENS (non-functioning)	43 (12)	65	73 (77.4)	30
Ketenci et al, Turkey	Physical therapy and rehabilitation	20-60	Thiocolchicoside tablets 8 mg	37	42		38
			Placebo oral (morning) & TZ (evening)	37	63		32
			Placebo oral (morning and evening)	40	52		27
Kofotolis et al, Greece	Not clear		Rhythmic stabilisation training	41 (6.4)	100	30 (6.2)	28
			Combination of isotonic exercises	42 (7.7)	100	33 (8.3)	28
			No treatment, instructed to avoid structured exercises or activities	42 (8.4)	100	35 (8.4)	30

Koumantakis et al, UK	Orthopaedic clinics + General practice		Stabilization enhanced exercise (Exercises of the stabilizing muscles in the back)	39 (11.4)		12	29
			General exercise (Exercises of the back extensors and flexors)	35 (9.7)		12	26
Kuukkanen et al, Finland	Occupational healthcare		Intensive training: intensive progressive exercises guided at the gym + home exercises		62		29
			Home exercise only: same as intensive, but unguided		48		29
			Control: usual activities, no trial exercises		54		28
Leclaire et al, Canada	Private physiatrist clinic	18-50	Standard care (rest, analgesics, physio) & Swedish back school	32 (7.7)	43		82
			Standard care (rest, analgesics, physio)	32 (8)	41		86
Lee et al, Korea	Not clear		Pulsed electromagnetic therapy (PEMT)	75 (5)	70	540 (661.5)	17
			Sham (Same machine, magnetic coil removed)	74 (4)	26	410 (499.5)	19
Licciardone et al, USA	General population	21-69	Osteopathic manipulative treatment	49 (12)	69		48
			Sham manipulation (Light touch & simulated osteopathic manipulation)	52 (12)	57		23
			No trial intervention (=usual back pain care)	49 (12)	65		20
Lindstrom et al, Sweden	Occupational healthcare		Swedish back school & workplace visit + graded exercise (CBT approach)		24		51
			Usual care: rest& analgesics & physical treatment		38		52
Linton et al, Sweden	General practice + general population	18-60	Back pain pamphlet	45	71		70
			Comprehensive information package	44	74		66
			CBT intervention	44	70		107
Long et al, Canada	Physical therapy	18-65	Matched direction exercise	43 (9.6)	51	14 (19.8)	80
			Opposite direction exercise	42 (10.3)	49	18 (21.8)	69
			Evidence based care	42 (10.8)	43	15 (17.6)	80
Malmivaara et al, Finland	Occupational healthcare		Complete bed rest (but for essential things)	40.8	60	0.7	67
			Back mobilising exercises	41.1	71	0.6	52
			Ordinary activities	39.1	70	0.6	67
Mannion et al 1999& 2001, Finland	General population	< 65	Modern active individual physiotherapy: strengthening, coordination and aerobics exercises, instructions on ergonomic principles + home exercises	46 (10.1)	61	520 (468)	46
			Muscle reconditioning on training devices (small groups)	45 (9.7)	54	504 (473.2)	47
			Low impact aerobic/stretching (groups)	44 (10.1)	55	676 (520)	44
Maul et al, Switzerland	Occupational healthcare	20-55	Back school & exercise	38 (8)			97
			Back school	39 (10)			86
Mayer et al, USA	General population + General practice	18-55	Heat wrap alone, 8h/d	29 (9.9)	68		25
			Exercise	33 (10.3)	60		25
			Heat wrap & exercise	32 (11.8)	63		24

			Booklet	31 (10.9)	92		26
Mehling et al, USA	General practice	20-70	Breath therapy	50 (12.1)	70	51 (26.6)	16
			Physical therapy: soft tissue mobilisation, joint mobilisation and exercises	49 (12.5)	58	57 (26.6)	12
Melancon et al, USA	Rural physician's practice	18-63	Massage therapy (Soft tissue massage)	39	46		30
			Traditional therapy (traditionally provided at US hospitals) = usual physicians care	38	45		30
Moffett et al, UK	General practice	18-60	Progressive exercise programme: group, Stretching, low impact aerobics & strengthening exercises, of all main muscles	41 (9.2)	57		85
			Usual GP care	43 (8.6)	56		98
Moseley L, Australia	Physiotherapy + General practice		Physiotherapy	43 (7)	64		29
			Usual GP care	38 (7)	54		28
Muehlbacher et al, Germany	General population	18 +	Topiramate tablets 50 mg	49 (5.4)	40	130	48
			Placebo tablets	47 (5)	35	104	48
Nadler et al, USA	Physiotherapy + general population	18-55	Wearable heat wrap	36 (11.6)			95
			Oral placebo tablets 2	37 (10.8)			96
			Ibuprofen tablets 400 mg	36 (11.6)			12
			Unheated wrap	35 (11.3)			16
Nadler et al, USA	General population	18-55	Wearable heat wrap	42 (9.4)	64		33
			Placebo tablets 2	42 (9.8)	62		34
			Ibuprofen tablets 400 mg	43 (2.7)	75		4
			Unheated wrap	34 (8.4)	80		5
Nadler et al, USA	Not clear	18-55	Continuous low level heat wrap therapy 8 hours/d	36 (10.5)	58		113
			Ibuprofen tablets 400 mg	35 (11.3)	57		113
			Acetaminophen tablets 500 mg, 2 tablets	37 (10.4)	60		106
			Placebo tablets, 2 tablets	37 (9.3)	58		20
			Unheated wrap	38 (9.1)	60		19
Niemisto et al 2003& 2005, Finland	General population	24-46	Manipulation, exercise & physician consultation	37 (5.6)	55	312	102
			Physician consultation only	37 (5.6)	53	312	102
Nordeman et al, Sweden	General practice + physical therapy dept		Early access to physio (Individualised, exercise, advice, group education)	39 (12.1)	63		32
			Waiting list control	41 (11.1)	50		28
Pallay et al, USA	General practice + specialists	18-75	Etoricoxib 90 mg tablets	53.3 (12.7)	59	634 (514.8)	106
			Etoricoxib 60 mg tablets	53.3 (12.3)	68	619 (629.2)	109
			Placebo tablets	51.8 (13.5)	61	598 (629.2)	110

Peloso et al, USA	Outpatients	18 +	Tramadol & Acetamenophen combination tablets 375/325 2 PRN	58 (11.5)	64		167
			Placebo tablets 2 PRN	58 (13.6)	61		169
Penttinen et al, Finland	Occupational healthcare	35-50	Back school & peer support		48		47
			Back school only		45		46
Perrot et al, France	Not clear	18 +	Tramadol / Paracetamol combination tablets 37.5/325 mg	57 (15.3)	64.4		59
			Tramadol tablets 50 mg	45 (14.6)	51.7		60
Pohjolainen et al, Finland	Hospital patients + Occupational healthcare		Oral Nimesulide & Placebo tablets* 100mg	42 (7.5)			52
			Ibuprofen tablets 600mg	42.5			52
Pope et al, USA	Health centre + general population	18-55	Spinal manipulation				70
			Soft tissue massage				37
			Transcutaneous muscle stimulation (=interferential therapy? 405)				28
			Corset				29
Preyde M, Canada	General practice + general population	18-81	Comprehensive massage: soft tissue manipulation, remedial exercises & Posture education)	48 (16.2)	56	12 (9.1)	25
			Soft tissue manipulation only	47 (18.4)	56	15 (8.2)	25
			Remedial exercise & posture education	48 (12.9)	41	13 (11.1)	25
			Sham laser therapy	42 (16.6)	54	13 (8.8)	25
Rasmussen et al, Sweden	Physiotherapy	18-60	Stabilizing training (Individual) (Cognitive + stabilisation of spinal muscles)	39 (12)	70		24
			Manual treatment (Individual) (Other muscles exercises, no manipulation)	37 (10)	78		23
Ruoff et al, USA	Not clear	25-75	Tramadol/ Acetamenophen combination tablets 37.5/325 mg, 2 tablets	54 (11.9)	67		161
			Placebo tablets 2 tablets	54 (12)	59		157
Rozenberg et al, France	Rheumatology clinics + General practice	18-65	Bed rest	44 (12.3)	49		137
			Normal activity	44 (11.8)	57		140
Rittweger et al, Germany	General population	40-60	Isodynamic lumbar extension	50 (6.6)	44	603 (520)	30
			Vibration exercise (On a machine with a vibrating platform)	54 (3.4)	52	754 (530.4)	30
Riipinen et al, Finland	General population	24-46	Manipulation exercise & physician consultation (combination)	39 (5.4)	55		102
			Physician consultation only	37 (6.5)	53		102
Ritvanen et al, Finland	General population	20-60	Traditional chiropractic bone setting	41 (4.9)	45		33
			Physical therapy	42 (6.8)	43		28
Rossignol et al, Canada	Workers compensation board		Coordination of primary healthcare program	37 (9.7)	33		54
			Usual GP care	38 (10.2)	23		56
Rossignol et al, France	General practice		Placebo tablets	40 (9.8)	58		81

			Atepadene tablets 90 mg	43 (9.9)	59		80
			The back book	44 (9.5)	59		76
			Atepadene & the back book	41 (10.9)	53		81
Rydeard et al,	Physicians clinics + physiotherapy + general population	22-55	Specific exercise treatment (Pilates, muscle activation)	37 (9)	57	286	21
			Waiting list control	34 (8)	72	468	18
Schnitzer et al, USA	Not clear	25-75	Tramadol tablets 100 mg				127
			Placebo tablets				127
Sherman et al, USA	Group health cooperative health insurance + general population	20-64	Yoga	44 (12)	68		36
			Exercise	42 (15)	63		35
			Self care book	44 (13)	67		30
Soukup et al, Norway	General practice + general population + referrals	18-50	Mensediesk exercise group intervention	40	53	676 (474.8)	34
			Waiting list group	40	49	578 (320.7)	35
Staal et al & Hlobil et al, Netherlands	Occupational healthcare		Graded activity (Physiotherapy + OT)	39 (9)	5	8.5	67
			Usual OT care	37 (8)	8	8	67
Stam et al, UK	General practice	18-65	Spiroflor SRL gel 3g	41 (13.6)	45		83
			Cremor Capsici Compositus FNA 3g	41 (12.8)	48		78
Steenstra et al, Netherlands	Occupational healthcare	18-65	Graded exercise program based on CBT	41 (9.2)	65		55
			Usual care by occupational healthcare therapist	43 (8.2)	54		57
Storheim et al, Norway	General practice + general population	20-60	Cognitive interaction	41 (9.4)	47		34
			Exercise (Back exercises and focus on ergonomic principles)	42 (9.2)	53		30
			Usual GP care	39 (11.9)	55		29
Suni et al, Finland	Occupational healthcare		Training group (exercises)	48 (5.8)	0		52
			Usual care (waiting list, received same treatment after 12 ms)	47 (5.3)	0		54
Szpalski et al, Belgium	Not clear		Tenoxicam, 20 mg od injection then tabs	38 (9.2)	38		37
			Placebo tablets, od, injection then tab	39 (10.4)	33		36
Thomas et al 2005& 2006, UK	General practice	18-65	Short course of acupuncture	42 (10.8)	62	17 (13.5)	159
			Usual GP care	44 (10.4)	58	17 (14.6)	80
Torstensen et al, Norway	Social security offices	20-65	Medical exercise therapy (MET)	42 (11.2)	52		71
			Conventional physiotherapy (CP)	43 (12)	48		67

			Self exercise	40 (11.4)	51		70
Triano et al, USA	Chiropractic clinic	18 +	High velocity low amplitude spinal manipulation				47
			High velocity low force mimic				39
			Back education programme				43
Tsui et al, Hong Kong	Physiotherapy	20-55	Electro-acupuncture & back exercise	40 (12.1)	76	39 (31.8)	14
			Electrical heat acupuncture + back exercise	39 (9.2)	71	54 (54.5)	14
			Back exercise only	41 (8.3)	62	50 (41.6)	14
Turner et al, USA	General practice + physicians + general population	20-65	Relaxation training (group)				24
			Cognitive therapy (group)				23
			Cognitive therapy & Relaxation training (group)				25
			Waiting list control				30
Underwood et al, UK	General practice	16-70	McKenzie technique + general advice	35	34	1	35
			General advice only	40	45	1	40
Wand et al, UK	General practice + A/E patients		Assess & Advice & Physiotherapy	34 (9)	44		43
			Assess & Advice & wait	35 (7.9)	55		51
Webster et al, USA	Not clear	18-70	Placebo tablets	48.7	61		101
			Oxycodone QDS	47.9	61		206
			Oxytrex QDS	47.8	62		206
			Oxytrex BD	47.9	62		206
Weiner et al, USA	General population	65 +	PENS & Physical therapy	74 (4.6)	65	551 (577.2)	17
			Sham PENS & Physical therapy	74 (5.7)	41	863 (852.8)	17
Werners et al, Germany	General practice		Interferential therapy: electrotherapy, to stimulate muscles fibres	38 (9.4)	43		68
			Motorised lumbar traction & massage	39 (9.5)	49		72
Williams et al, USA	Physicians + general population	18 +	Education (Lectures in physio/OT and written hand-outs on back pain)	48 (10.7)	71		30
			Yoga	49 (5.8)	65		30
Wright et al, UK	Occupational healthcare + General practice		The back book & verbal advice + usual GP care	44 (11.5)	60		56
			The back book & verbal advice & single manipulation or injection & individual and group exercise	38 (13.5)	47		55
Yelland et al, Australia	General practice	21-70	Glucose lignocaine injection	52 (10.6)	59	769.6 (556.8)	28
			Exercise (Alternating: flexion and extension of spine and hips)	49 (10.4)	55	718 (483.6)	26
			Saline injection	50 (9.8)	56	759 (499.2)	27

		Normal activity	51 (11.2)	58	733 (551.2)	29
Yokoyama M, Japan	Not clear	PENS: percutaneous electrical nerve stimulation	60 (12)	61	15 (7)	17
		PENS & TENS	58 (14)	53	15 (8)	17
		TENS	50 (13)	56	13 (6)	18
Zerbini C, Multinational	Not clear	Etoricoxib tablets 60 mg	52 (14.7)	72	432 (457.6)	224
		Diclofenac tablets 50 mg	52 (12.8)	72	432 (468)	222

Appendix 4.2b Citations of papers (*n* 126) of trials (*n* 118) included in the systematic review presented in Chapter Four.

Andersson 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 Eng J Med*; 341:1426-1431.

Bannwarth B, Allaert FA, Avouac B, Rossignol M, Rozenberg S, Valat JP (2005). A randomized, double-blind, placebo controlled study of oral adenosine triphosphate in subacute low back pain. *J Rheumatology*; 32:1114-1117.

Bendix T, Bendix A, Labriola M, Haestrup C, Ebbehøj N (2000). Functional restoration versus outpatient physical training in chronic low back pain: a randomized comparative study. *Spine*; 25:2494-2500.

Beurskens AJ, de-Vet HC, Köke AJ, Lindeman E, Regtop W, van-der-Heijden GJ et al (1995). Efficacy of traction for non-specific low back pain: a randomised clinical trial. *Lancet*; 346:1596-1600.

Birbara CA, Puopolo AD, Munoz DR, Sheldon EA, Mangione A, Bohidar NR et al (2003). Treatment of chronic low back pain with etoricoxib, a new cyclo-oxygenase-2 selective inhibitor: improvement in pain and disability--a randomized, placebo-controlled, 3-month trial. *J Pain*; 4:307-315.

Blomberg S, Svärdsudd K, Tibblin G (1994). A randomized study of manual therapy with steroid injections in low-back pain. Telephone interview follow-up of pain, disability, recovery and drug consumption. *Eur Spine J*; 3:246-254.

Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE (2006). Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. *Spine*; 31:623-631.

Bronfort G, Goldsmith CH, Nelson CF, Boline PD, Anderson AV (1996). Trunk exercise combined with spinal manipulative or NSAID therapy for chronic low back pain: a randomized, observer-blinded clinical trial. *J Manip Physio therapeutics*;19:570-582.

Burton AK, Waddell G, Tillotson KM, Summerton N (1999). Information and advice to patients with back pain can have a positive effect. A randomized controlled trial of a novel educational booklet in primary care. *Spine*; 24:2484-2491.

Cambron JA, Gudavalli MR, Hedeker D, McGregor M, Jedlicka J, Keenum M et Al (2006). One-year follow-up of a randomized clinical trial comparing flexion distraction with an exercise program for chronic low-back pain. *J Alt Comp Med*; 12:659-668.

Carr JL, Klaber-Moffett JA, Howarth E, Richmond SJ, Torgerson DJ, Jackson DA et al (2005). A randomized trial comparing a group exercise programme for back pain patients with individual physiotherapy in a severely deprived area. *Dis Rehab*; 27:929-937.

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. *N Eng J Med*; 339:1021-1029.

Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR et al (2004). A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Int Med*; 141:920-928.

Chiradejnant A, Maher CG, Latimer J, Stepkovitch N (2003). Efficacy of "therapist-selected" versus "randomly selected" mobilisation techniques for the treatment of low back pain: a randomised controlled trial. *Aust J Physio*; 49:233-241.

Chok B, Lee R, Latimer J, Tan SB (1999). Endurance training of the trunk extensor muscles in people with subacute low back pain. *Phys Ther*; 79:1032-1042.

Cleland JA, Childs JD, Palmer JA, Eberhart S (2006). Slump stretching in the management of non-radicular low back pain: a pilot clinical trial. *Man Ther*; 11:279-286.

Coats TL, Borenstein DG, Nangia NK, Brown MT (2004). Effects of valdecoxib in the treatment of chronic low back pain: results of a randomized, placebo-controlled trial. *Clin Ther*; 26:1249-1260.

Constant F, Collin JF, Guillemin F, Boulangé M (1995). Effectiveness of spa therapy in chronic low back pain: a randomized clinical trial. *J Rheumatology*; 22:1315-1320.

Cramer GD, Humphreys CR, Hondras MA, McGregor M, Triano JJ (1993). The Hmax/Mmax ratio as an outcome measure for acute low back pain. *J Man Physio Ther*; 16:7-13.

Damush TM, Weinberger M, Perkins SM, Rao JK, Tierney WM, Qi R et al (2004). Randomized trial of a self-management program for primary care patients with acute low back pain: short-term effects. *Arthr Rheumatism*; 49:179-186.

Damush TM, Weinberger M, Perkins SM, Rao JK, Tierney WM, Qi R et al (2003). The long-term effects of a self-management program for inner-city primary care patients with acute low back pain. *Arch Int Med*; 163:2632-2638.

Defrin R, Ben BS, Aldubi RD, Pick CG (2005). Conservative correction of leg-length discrepancies of 10mm or less for the relief of chronic low back pain. *Arch Phys Med Rehab*; 86:2075-2080.

Delitto A, Cibulka MT, Erhard RE, Bowling RW, Tenhula JA (1993). Evidence for use of an extension-mobilization category in acute low back syndrome: a prescriptive validation pilot study. *Phys Ther*; 73:216-222.

Descarreaux M, Normand MC, Laurencelle L, Dugas C (2002). Evaluation of a specific home exercise program for low back pain. *J Manip Physio Ther*; 25:497-503.

Dreiser RL, Marty M, Ionescu E, Gold M, Liu JH (2003). Relief of acute low back pain with diclofenac-K 12.5 mg tablets: a flexible dose, ibuprofen 200 mg and placebo-controlled clinical trial. *Int J Clin Pharm Ther*; 41:375-385.

Faas A, Chavannes AW, van-Eijk JT, Gubbels JW (1993). A randomized, placebo-controlled trial of exercise therapy in patients with acute low back pain. *Spine*; 18:1388-1395.

Frost H, Lamb SE, Doll HA, Carver PT, Stewart BS (2004). Randomised controlled trial of physiotherapy compared with advice for low back pain. *BMJ*; 329:708.

Geisser ME, Wiggert EA, Haig AJ, Colwell MO (2005). A randomized, controlled trial of manual therapy and specific adjuvant exercise for chronic low back pain. *The Clin J Pain*; 21:463-470.

Gemmell HA, Jacobson BH (1995). The immediate effect of activator vs. meric adjustment on acute low back pain: a randomized controlled trial. *J Manip Physio Ther*; 18:453-456.

George SZ, Fritz JM, Bialosky JE, Donald DA (2003). The effect of a fear-avoidance-based physical therapy intervention for patients with acute low back pain: results of a randomized clinical trial. *Spine*; 28:2551-2560.

Giles LG, Muller R (2003). Chronic spinal pain: a randomized clinical trial comparing medication, acupuncture, and spinal manipulation. *Spine*; 28:1490-1502.

Glomsrød B, Lønn JH, Soukup MG, Bø K, Larsen S (2001). "Active back school", prophylactic management for low back pain: three-year follow-up of a randomized, controlled trial. *J Rehab Med*; 33:26-30.

Goldby LJ, Moore AP, Doust J, Trew ME (2006). A randomized controlled trial investigating the efficiency of musculoskeletal physiotherapy on chronic low back

disorder. *Spine*; 31:1083-1093.

Grunnesjö MI, Bogefeldt JP, Svärdsudd KF, Blomberg SI (2004). A randomized controlled clinical trial of stay-active care versus manual therapy in addition to stay-active care: functional variables and pain. *J Manip Physio Ther* 2004; 27: 431-441.

Gudavalli MR, Cambron JA, McGregor M, Jedlicka J, Keenum M, Ghanayem AJ et al (2006). A randomized clinical trial and subgroup analysis to compare flexion-distraction with active exercise for chronic low back pain. *Eur Spin J*; 15:1070-1082.

Guillemin F, Constant F, Collin JF, Boulange M (1994). Short and long-term effect of spa therapy in chronic low back pain. *BJ Rheumatology*; 33:148-151.

Haas M, Groupp E, MuenchJ, KraemerD, BrummelSmith K, SharmaR, Ganger B, Attwood M, Fairweather A (2005). Chronic disease self-management program for low back pain in the elderly. *J Manip Physio Ther*; 28: 228-237.

Hansen FR, Bendix T, Skov P, Jensen CV, Kristensen JH, Krohn L et al (1993). Intensive, dynamic back-muscle exercises, conventional physiotherapy, or placebo-control treatment of low-back pain. A randomised, observer-blind trial. *Spine*; 18:98-108.

Hay EM, Mullis R, Lewis M, Vohora K, Main CJ, Watson P et al (2005). Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet*; 365:2024-2030.

Hawk C, Long CR, Rowell RM, Gudavalli MR, Jedlicka J (2005). A randomized trial investigating a chiropractic manual placebo: a novel design using standardized forces in the delivery of active and control treatments. *J Alt Comp Med*; 11:109-117.

Helmhout PH, Harts CC, Staal JB, Candel MJ, de-Bie RA (2004). Comparison of a high-intensity and a low-intensity lumbar extensor training program as minimal intervention treatment in low back pain: a randomized trial. *Eur Spine J*;13:537-547.

Hemmilä HM, Keinänen-Kiukaanniemi SM, Levoska S, Puska P (2002). Long-term effectiveness of bone-setting, light exercise therapy, and physiotherapy for prolonged back pain: a randomized controlled trial. *J Manip Physio Ther*; 25:99-104.

Heymans MW, de-Vet HC, Bongers PM, Knol DL, Koes BW, van Mechelen W (2006). The effectiveness of high-intensity versus low-intensity back schools in

an occupational setting: a pragmatic randomized controlled trial. *Spine*; 31:1075-1082.

Hlobil H, Staal JB, Twisk J, Köke A, Ariëns G, Smid T et al (2005). The effects of a graded activity intervention for low back pain in occupational health on sick leave, functional status and pain: 12-month results of a randomized controlled trial. *J Occup Rehab*; 15:569-580.

Hoiriis KT, Pfleger B, McDuffie FC, Cotsonis G, Elsangak O, Hinson R et al (2004). A randomized clinical trial comparing chiropractic adjustments to muscle relaxants for subacute low back pain. *J Manip Physiol Ther*; 27:388-398.

Hsieh CY, Adams AH, Tobis J, Hong CZ, Danielson C, Platt K et al (2002). Effectiveness of four conservative treatments for subacute low back pain: a randomized clinical trial. *Spine*; 27:1142-1148.

Hurley DA, McDonough SM, Dempster M, Moore AP, Baxter GD (2004). A randomized clinical trial of manipulative therapy and interferential therapy for acute low back pain. *Spine*; 29:2207-2216.

Hurley DA, Minder PM, McDonough SM, Walsh DM, Moore AP, Baxter DG (2001). Interferential therapy electrode placement technique in acute low back pain: a preliminary investigation. *Arch Phys Med Rehab*; 82:485-493.

Hurwitz EL, Morgenstern H, Harber P, Kominski GF, Belin TR, Yu F et al (2002). A randomized trial of medical care with and without physical therapy and chiropractic care with and without physical modalities for patients with low back pain: 6-month follow-up outcomes from the UCLA low back pain study. *Spine*; 27:2193-2204.

Hurwitz EL, Morgenstern H, Harber P, Kominski GF, Belin TR, Yu F et al (2002). Second Prize: The effectiveness of physical modalities among patients with low back pain randomized to chiropractic care: findings from the UCLA low back pain study. *J Manip Physiol Ther*; 25:10-20.

Jellema P, van-der-Windt DA, van-der-Horst HE, Twisk JW, Stalman WA, Bouter LM (2005). Should treatment of (sub)acute low back pain be aimed at psychosocial prognostic factors? Cluster randomised clinical trial in general practice. *BMJ*; 331:84.

Kääpä EH, Frantsi K, Sarna S, Malmivaara A (2006). Multidisciplinary group rehabilitation versus individual physiotherapy for chronic nonspecific low back pain: a randomized trial. *Spine*; 31:371-376.

Kankaanpää M, Taimela S, Airaksinen O, Hänninen O (1999). The efficacy of active rehabilitation in chronic low back pain. Effect on pain intensity, self-

experienced disability, and lumbar fatigability. *Spine*; 24:1034-1042.

Karjalainen K, Malmivaara A, Mutanen P, Roine R, Hurri H, Pohjolainen T (2004). Mini-intervention for subacute low back pain: two-year follow-up and modifiers of effectiveness. *Spine*; 29:1069-1076.

Karjalainen K, Malmivaara A, Pohjolainen T, Hurri H, Mutanen P, Rissanen P et Al (2003). Mini-intervention for subacute low back pain: a randomized controlled trial. *Spine*; 28:533-540.

Katz N, Rodgers DB, Krupa D, Reicin A (2004). Onset of pain relief with rofecoxib in chronic low back pain: results of two four-week, randomized, placebo-controlled trials. *Current medical research and opinion*; 20:651-658.

Katz N, Ju WD, Krupa DA, Sperling RS, Bozalis RD, Gertz BJ et al (2003). Efficacy and safety of rofecoxib in patients with chronic low back pain: results from two 4-week, randomized, placebo-controlled, parallel-group, double-blind trials. *Spine*; 28:851-858.

Katz J, Pennella VJ, Hetzel RD, Kanazi GE, Dworkin RH (2005). A randomized, placebo-controlled trial of bupropion sustained release in chronic low back pain. *J Pain* 2005; 6:656-661.

Kerr DP, Walsh DM, Baxter D (2003). Acupuncture in the management of chronic low back pain: a blinded randomized controlled trial. *Clin J Pain*; 19:364-370.

Ketenci A, Ozcan E, Karamursel S (2005). Assessment of efficacy and psychomotor performances of thiocolchicoside and tizanidine in patients with acute low back pain. *Inter J Clin Pract;* 59:764-770.

Kofotolis N, Kellis E (2006) Effects of two 4-week proprioceptive neuromuscular facilitation programs on muscle endurance, flexibility, and functional performance in women with chronic low back pain. *Phys Ther*; 86:1001-1012.

Koumantakis GA, Watson PJ, Oldham JA (2005). Trunk muscle stabilization training plus general exercise versus general exercise only: randomized controlled trial of patients with recurrent low back pain. *Phys Ther*; 85:209-225.

Kuukkanen T, Mälkiä E (2000). Effects of a three-month therapeutic exercise programme on flexibility in subjects with low back pain. *Physio Res Inter*; 5:46-61.

Leclaire R, Esdaile JM, Suissa S, Rossignol M, Proulx R, Dupuis M (1996). Back school in a first episode of compensated acute low back pain: a clinical trial to assess efficacy and prevent relapse*. Arch Phys Med Rehab*; 77:673-679.

Lee PB, Kim YC, Lim YJ, Lee CJ, Choi SS, Park SH et al (2006). Efficacy of pulsed electromagnetic therapy for chronic lower back pain: a randomized, double-blind, placebo-controlled study. *J Inter Med Res*; 34:160-167.

Licciardone JC, Stoll ST, Fulda KG, Russo DP, Siu J, Winn W et al (2003). Osteopathic manipulative treatment for chronic low back pain: a randomised controlled trial. *Spine*; 28:1355-1362.

Lindström I, Ohlund C, Nachemson A (1995). Physical performance, pain, pain behavior and subjective disability in patients with subacute low back pain. *Scand J Rehab Med*; 27:153-160.

Linton SJ, Andersson T (2000). Can chronic disability be prevented? A randomised trial of a cognitive-behavior intervention and two forms of information for patients with spinal pain. *Spine*; 25:2825-2831.

Long A, Donelson R, Fung T (2004). Does it matter which exercise? A Randomized control trial of exercise for low back pain. *Spine*; 29:2593-2602.

Malmivaara A, Häkkinen U, Aro T, Heinrichs ML, Koskenniemi L, Kuosma E et al (1995). The treatment of acute low back pain--bed rest, exercises, or ordinary activity? *N Eng J Med*; 332:351-355.

Mannion AF, Müntener M, Taimela S, Dvorak J (1999). A randomized clinical trial of three active therapies for chronic low back pain. *Spine*; 24:2435-2448.

Mannion AF, Müntener M, Taimela S, Dvorak J (2001) Comparison of three active therapies for chronic low back pain: results of a randomized clinical trial with one-year follow-up. *Rheumatology*; 40:772-778.

Maul I, Läubli T, Oliveri M, Krueger H (2005). Long-term effects of supervised physical training in secondary prevention of low back pain. *Eur Spine J*; 14:599-611.

Mayer JM, Ralph L, Look M, Erasala GN, Verna JL, Matheson LN et al (2005). Treating acute low back pain with continuous low-level heat wrap therapy and/or exercise: a randomized controlled trial. *Spine J*; 5:395-403.

Mehling WE, Hamel KA, Acree M, Byl N, Hecht FM (2005). Randomized, controlled trial of breath therapy for patients with chronic low-back pain. *Alt Ther Health Med*; 11:44-52.

Melancon B, Miller LH (2005). Massage therapy versus traditional therapy for low back pain relief: implications for holistic nursing practice. *Hol Nursing pract*; 19:116-121.

Moffett JK, Torgerson D, Bell SS, Jackson D, Llewlyn PH, Farrin A et al (1999). Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. *BMJ*; 319:279-283.

Moseley L (2002). Combined physiotherapy and education is efficacious for chronic low back pain. *Aust J Physio* 2002; 48:297-302.

Muehlbacher M, Nickel MK, Kettler C, Tritt K, Lahmann C, Leiberich PK et al (2006). Topiramate in treatment of patients with chronic low back pain: a randomised, double-blind, placebo-controlled study. *Clin J Pain*; 22:526-531.

Nadler SF, Steiner DJ, Erasala GN, Hengehold DA, Abeln SB, Weingand KW (2003). Continuous low-level heat wrap therapy for treating acute nonspecific low back pain. *Arch Phys Med Rehab*; 84:329-334.

Nadler SF, Steiner DJ, Petty SR, Erasala GN, Hengehold DA, Weingand KW (2003). Overnight use of continuous low-level heat wrap therapy for relief of low back pain. *Arch Phys Med Rehab*; 84:335-342.

Nadler SF, Steiner DJ, Erasala GN, Hengehold DA, Hinkle RT, Beth GM et al (2002). Continuous low-level heat wrap therapy provides more efficacy than Ibuprofen and acetaminophen for acute low back pain. *Spine*; 27:1012-1017.

Niemistö L, Lahtinen ST, Rissanen P, Lindgren KA, Sarna S, Hurri H (2003). A randomized trial of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain. *Spine*; 28:2185-2191.

Niemistö L, Rissanen P, Sarna S, Lahtinen ST, Lindgren KA, Hurri H (2005). Cost-effectiveness of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain: a prospective randomized trial with 2-year follow-up. *Spine*; 30:1109 1115.

Nordeman L, Nilsson B, Möller M, Gunnarsson R (2006). Early access to physical therapy treatment for subacute low back pain in primary health care: a prospective randomized clinical trial. *The Clin J Pain*; 22:505-511.

Pallay RM, Seger W, Adler JL, Ettlinger RE, Quaidoo EA, Lipetz R et al (2004). Etoricoxib reduced pain and disability and improved quality of life in patients with chronic low back pain: a 3 month, randomized, controlled trial. *Scand J Rheum*; 33:257-266.

Peloso PM, Fortin L, Beaulieu A, Kamin M, Rosenthal N (2004). Analgesic efficacy and safety of tramadol/ acetaminophen combination tablets (Ultracet) in

treatment of chronic low back pain: a multicenter, outpatient, randomized, double blind, placebo controlled trial. *J Rheum* ; 31:2454-2463.

Penttinen J, Nevala PN, Airaksinen O, Jääskeläinen M, Sintonen H, Takala J (2002). Randomized controlled trial of back school with and without peer support. *J Occup Rehab*; 12:21-29.

Perrot S, Krause D, Crozes P, Naïm C (2006). Efficacy and tolerability of paracetamol/tramadol (325 mg/37.5 mg) combination treatment compared with tramadol (50 mg) monotherapy in patients with subacute low back pain: a multicenter, randomized, double-blind, parallel-group, 10-day treatment study. *Clin Ther*; 28:1592-1606.

Pohjolainen T, Jekunen A, Autio L, Vuorela H (2000). Treatment of acute low back pain with the COX-2-selective anti-inflammatory drug nimesulide: results of a randomized, double-blind comparative trial versus ibuprofen. *Spine*; 25:1579-1585.

Pope MH, Phillips RB, Haugh LD, Hsieh CY, MacDonald L, Haldeman S (1994). A prospective randomized three-week trial of spinal manipulation, transcutaneous muscle stimulation, massage and corset in the treatment of subacute low back pain. *Spine*; 19:2571-2577.

Preyde M (2000). Effectiveness of massage therapy for subacute low-back pain: a randomized controlled trial. *CMAJ*; 162:1815-1820.

Rasmussen BE, Nilsson WL, Arvidsson I (2003). Stabilizing training compared with manual treatment in sub-acute and chronic low-back pain. *Man Ther*; 8:233-241.

Rozenberg S, Delval C, Rezvani Y, Olivieri AN, Kuntz JL, Legrand E et al (2002). Bed rest or normal activity for patients with acute low back pain: a randomized controlled trial. *Spine*; 27:1487-1493.

Riipinen M, Niemistö L, Lindgren KA, Hurri H (2005). Psychosocial differences as predictors for recovery from chronic low back pain following manipulation, stabilizing exercises and physician consultation or physician consultation alone. *J Rehab Med*; 37:152-158.

Rittweger J, Just K, Kautzsch K, Reeg P, Felsenberg D (2002). Treatment of chronic lower back pain with lumbar extension and whole-body vibration exercise: a randomized controlled trial. *Spine*; 27:1829-1834.

Ritvanen T, Zaproudina N, Nissen M, Leinonen V, Hänninen O (2007). Dynamic Surface electromyographic responses in chronic low back pain treated by traditional bone setting and conventional physical therapy. *J Manip Physio Ther*;

30:31-37.

Rossignol M, Abenhaim L, Séguin P, Neveu A, Collet JP, Ducruet T et al (2000). Coordination of primary health care for back pain. A randomised controlled trial. *Spine*; 25:251-258.

Rossignol M, Allaert FA, Rozenberg S, Valat JP, Avouac B, Peres G et al (2005). Measuring the contribution of pharmacological treatment to advice to stay active in patients with subacute low-back pain: a randomised controlled trial. *Pharmacoepidemiology and drug safety*; 14:861-867.

Ruoff GE, Rosenthal N, Jordan D, Karim R, Kamin M (2003). Tramadol/acetaminophen combination tablets for the treatment of chronic lower back pain: a multicenter, randomized, double-blind, placebo-controlled outpatient study. *Clin ther*; 25:1123-1141.

Rydeard R, Leger A, Smith D (2006). Pilates-based therapeutic exercise: effect on subjects with nonspecific chronic low back pain and functional disability: a randomized controlled trial. *J Ortho Sports Phys Ther*; 36:472-484.

Schnitzer TJ, Gray WL, Paster RZ, Kamin M (2000). Efficacy of tramadol in treatment of chronic low back pain. *J Rheum*; 27:772-778.

Sherman KJ, Cherkin DC, Erro J, Miglioretti DL, Deyo RA (2005). Comparing yoga, exercise, and a self-care book for chronic low back pain: a randomized, controlled trial. *Ann Int Med*; 143:849-856.

Soukup MG, Glomsröd B, Lönn JH, Bö K, Larsen S (1999). The effect of a Mensendieck exercise program as secondary prophylaxis for recurrent low back pain. A randomized, controlled trial with 12-month follow-up. *Spine*; 24:1585 1591.

Staal JB, Hlobil H, Twisk JW, Smid T, Köke AJ, van Mechelen W (2004). Graded activity for low back pain in occupational health care: a randomized, controlled trial. *Ann Int Med*; 140:77-84.

Stam C, Bonnet MS, van-Haselen RA (2001). The efficacy and safety of a homeopathic gel in the treatment of acute low back pain: a multi-centre, randomised, double-blind comparative clinical trial. *Br Homoeo J*; 90:21-28.

Steenstra IA, Anema JR, Bongers PM, de-Vet HC, Knol DL, van Mechelen W (2006). The effectiveness of graded activity for low back pain in occupational healthcare. *Occu Envir Med*; 63:718-725.

Storheim K, Brox JI, Holm I, Koller AK, Bø K (2003). Intensive group training Versus cognitive intervention in sub-acute low back pain: short-term results of a

single-blind randomized controlled trial. *J Rehab Med*; 35:132-140.

Suni J, Rinne M, Natri A, Statistisian MP, Parkkari J, Alaranta H (2006). Control of the lumbar neutral zone decreases low back pain and improves self-evaluated work ability: a 12-month randomized controlled study. *Spine*; 31:E611-E620.

Szpalski M, Hayez JP (1994). Objective functional assessment of the efficacy of tenoxicam in the treatment of acute low back pain. A double-blind placebo-controlled study. *BJ Rheum*; 33:74-78.

Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M et al (2005). Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain. *Health Technology Assessment*; 9:iii-x, 1.

Thomas KJ, MacPherson H, Thorpe L, Brazier J, Fitter M, Campbell MJ et al (2006). Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. *BMJ*; 333:623.

Torstensen TA, Ljunggren AE, Meen HD, Odland E, Mowinckel P, Geijerstam S (1998). Efficiency and costs of medical exercise therapy, conventional physiotherapy, and self-exercise in patients with chronic low back pain. A pragmatic, randomized, single-blinded, controlled trial with 1-year follow-up. *Spine*; 23:2616-2624.

Triano JJ, McGregor M, Hondras MA, Brennan PC (1995). Manipulative therapy versus education programs in chronic low back pain. *Spine*; 20:948-955.

Tsui ML, Cheing GL (2004). The effectiveness of electroacupuncture versus electrical heat acupuncture in the management of chronic low-back pain. *J Alt Comp Med*; 10:803-809.

Turner JA, Jensen MP (1993). Efficacy of cognitive therapy for chronic low back pain. *Pain*; 52:169-177.

Underwood MR, Morgan J (1998). The use of a back class teaching extension exercises in the treatment of acute low back pain in primary care. *Family Practice*; 15:9-15.

Wand BM, Bird C, McAuley JH, Doré CJ, MacDowell M, De-Souza LH (2004). Early intervention for the management of acute low back pain: a single-blind randomized controlled trial of biopsychosocial education, manual therapy, and exercise. *Spine*; 29:2350-2356.

Webster LR, Butera PG, Moran LV, Wu N, Burns LH, Friedmann N (2006). Oxytrex minimizes physical dependence while providing effective analgesia: a

randomised controlled trial in low back pain. *J Pain*; 7:937-946.

Weiner DK, Rudy TE, Glick RM, Boston JR, Lieber SJ, Morrow LA et al (2003). Efficacy of percutaneous electrical nerve stimulation for the treatment of chronic low back pain in older adults. *J Am Ger Soc*; 51:599-608.

Werners R, Pynsent PB, Bulstrode CJ (1999). Randomized trial comparing interferential therapy with motorized lumbar traction and massage in the management of low back pain in a primary care setting. *Spine*; 24:1579-1584.

Williams KA, Petronis J, Smith D, Goodrich D, Wu J, Ravi N et al (2005). Effect of Iyengar yoga therapy for chronic low back pain. *Pain*; 115:107-117.

Wright A, Lloyd DA, Williams S, Ellis R, Strike P (2005). Individual active treatment combined with group exercise for acute and subacute low back pain. *Spine* 2005; 30:1235-1241.

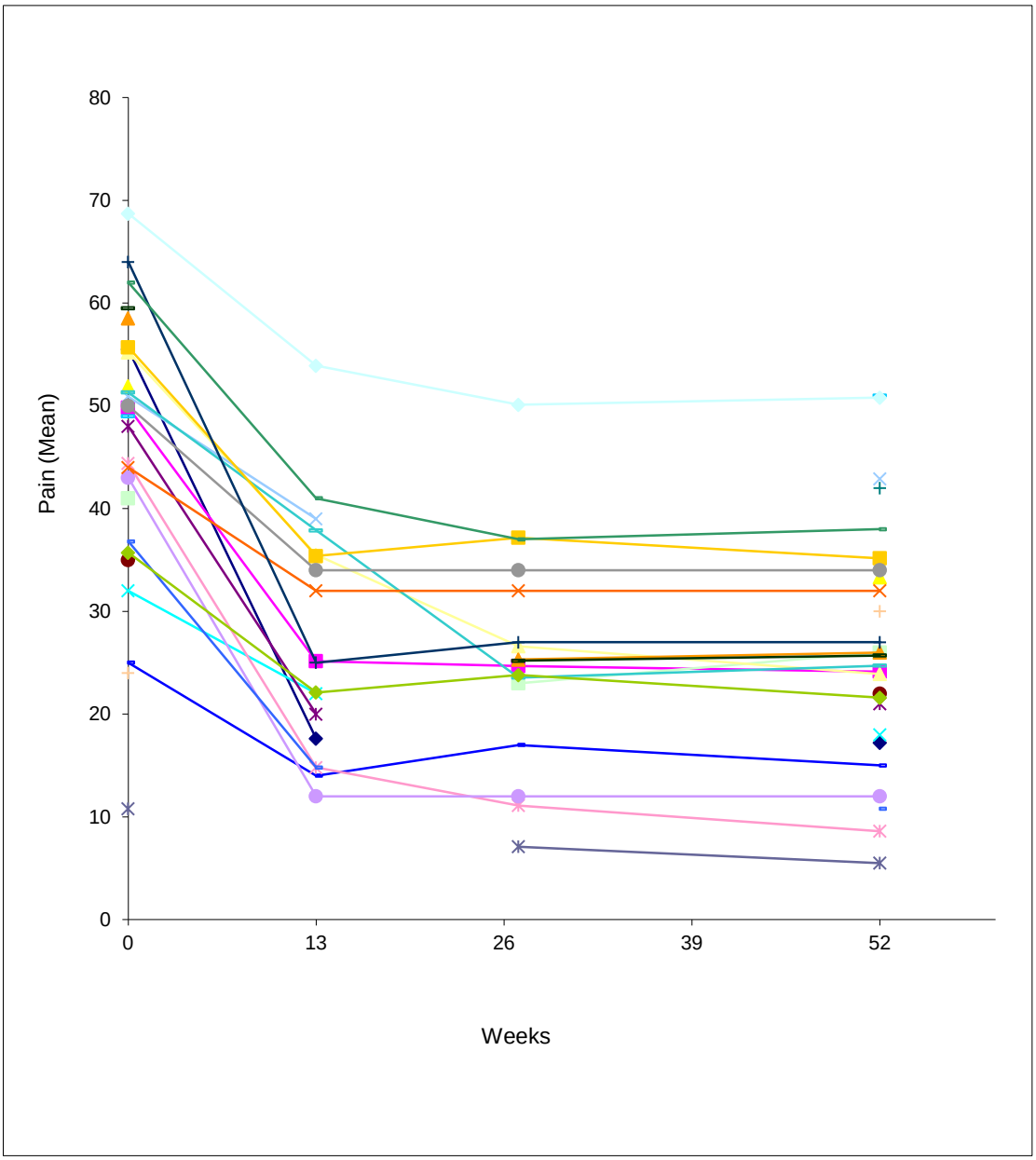
Yelland MJ, Glasziou PP, Bogduk N, Schluter PJ, McKernon M (2004). Prolotherapy injections, saline injections, and exercises for chronic low-back pain: a randomized trial. *Spine*; 29:9-16.

Yokoyama M, Sun X, Oku S, Taga N, Sato K, Mizobuchi S et al (2004). Comparison of percutaneous electrical nerve stimulation with transcutaneous electrical nerve stimulation for long-term pain relief in patients with chronic low back pain. *Anesth Analg*; 98:1552-6

Zerbini C, Ozturk ZE, Grifka J, Maini M, Nilganuwong S, Morales R et al (2005). Efficacy of etoricoxib 60 mg/day and diclofenac 150 mg/day in reduction of pain and disability in patients with chronic low back pain: results of a 4-week, multinational, randomised, double-blind study. *Cur Med Res Op*; 21:2037-2049.

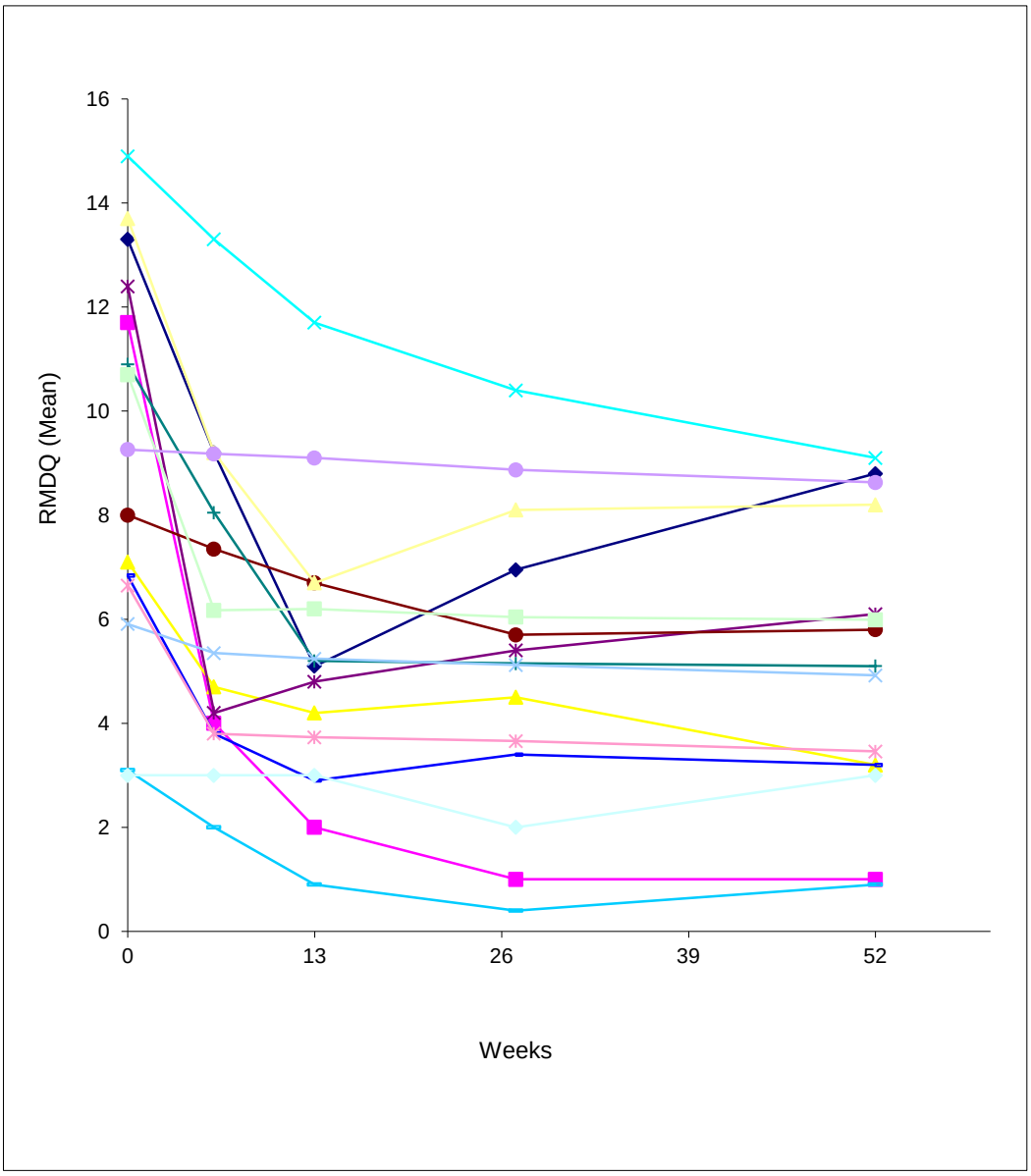
Appendix 4.3a Overall symptom progression (VAS for pain) up to 52 week follow up in one randomly selected treatment arm from each trial.

Each line represents a response line within a trial arm.



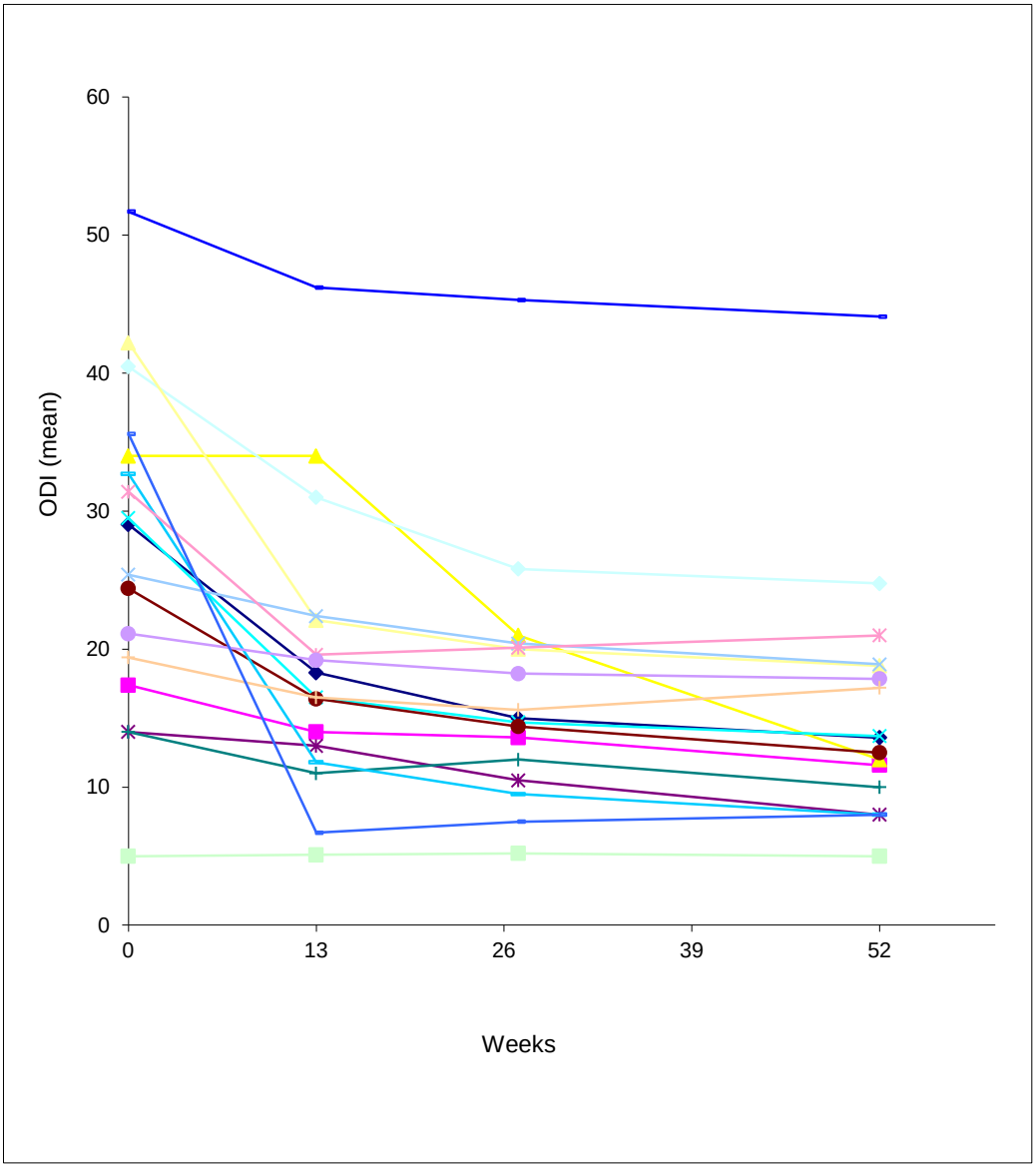
Appendix 4.3b Overall symptom progression (RMDQ) up to 52 week follow up in one randomly selected treatment arm from each trial

Each line represents a response line within a trial arm.



Appendix 4.3c Overall symptom progression (ODI) up to 52 week follow up in one randomly selected treatment arm from each trial

Each line represents a response line within a trial arm.



Appendix 5 Trials' quality assessment criteria used for Cochrane systematic reviews and the modifications that have been introduced for use in Chapter Five.

Original	Modified
1. Was the method of randomization adequate?	1m. Was the method of randomization adequate?
2. Was the treatment allocation concealed?	2m. Was the treatment allocation concealed?
3. Were the groups similar at baseline regarding the most important prognostic factors?	--------
4. Was the patient blinded to the intervention?	3m. Was the patient blinded to the intervention?
5. Was the care provider blinded to the intervention?	4m. Was the care provider blinded to the intervention?
6. Was the outcome assessor blinded to the intervention?	--------
7. Were co-interventions avoided or similar?	5m. Were co-interventions prevented/avoided?
	6m. Were co-interventions standardised?
8 Was the compliance acceptable in all groups?	7m. Was compliance acceptable?
9. Was the drop-out rate described and acceptable?	8m. Was the drop-out rate acceptable?
10. Was the timing of the outcome assessment in both groups comparable?	9m. Was the timing of the outcome assessment in all groups comparable?
11. Did the study include an intention to treat analysis?	10m. Was the analysis based on intention-to-treat analysis?
Instructions for completion	***Instructions for completion***
1. A random (unpredictable) assignment sequence. Examples of adequate methods are computer generated random number table and use of sealed opaque envelopes. Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should not be regarded as appropriate.	1m. For cluster trials only: Examples of adequate methods are computer generated random number table and use of sealed opaque envelopes. Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should not be regarded as appropriate.
2. Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.	2m. Same as point (1) for all other random trials.
3. In order to receive a 'yes', groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurological symptoms and value of main outcome measure(s).	3m. Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.
4. The reviewer determines if there was enough information about the blinding of the patient to score a "yes".	-- -- -- --
	4m. The reviewer determines if there was enough information about the blinding of the patient to score a "yes".
5. The reviewer determines if there was enough information about the blinding of the care provider to score a "yes".	5m. The reviewer determines if there was enough information about the blinding of the care provider to score a "yes".
6. The reviewer determines if there was enough information about the blinding of the outcome assessor to score a "yes".	-- -- -- --
	6m. The reviewer determines if co-interventions were avoided in the trial design to score a "yes".
7. Co-interventions should either be avoided in the trial design or comparable between the index and control groups.	7m. The reviewer determines if co-interventions were comparable between the index and control groups and standardised to score a "yes". If item 6 is yes, then item 7 is not applicable.
-- -- -- --	
8. The reviewer determines if the compliance to the interventions is acceptable, based on the reported intensity, duration, number, and frequency of sessions for both the index intervention(s) and control intervention(s).	8m. The reviewer determines if the compliance to the interventions is acceptable, based on the reported intensity, duration, number, and frequency of sessions for both the index intervention(s) and control intervention(s).
9. The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias, a "yes" is scored. (N.B., these percentages	9m. The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias, a "yes" is scored. *(N.B., these percentages*

are arbitrary, not supported by literature). 10. Timing of outcome assessment should be identical for all intervention groups and for all important outcome assessments. 11. All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of non-compliance and co-interventions.	*are arbitrary, not supported by literature).* 10m. Timing of outcome assessment should be identical for all groups and for all important outcome assessments. 11m. All randomised patients are reported / analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of non-compliance and co-interventions.

Appendix 6 Summary of trials (n 44) that provided data for pain intensity outcome and were included in the meta-regression analyses presented in Chapter Six. Full citations of included trials are presented in Appendix 4.2b

Author and country	Setting	Age (y)	Treatments	Age (y) mean (SD)	Female (%)	Actual duration of LBP (weeks) mean (SD)	Sample of trial arm at baseline
Bendix et al, Denmark	General practice		Functional restoration (PT + OT + Psychological)	40	66		48
			Outpatient intensive physical training: Aerobics + strengthening exercises + fitness machines	43	69		51
Burton et al, UK	General practice	17-70	The Back Book + usual care (GP & osteopathic care)		11		83
			The traditional Handy Hints & usual care (GP & osteopathic care)		12		79
Cambron et al, USA	Chiropractic clinic + hospital clinic + general population	18 +	Chiropractic flexion distraction procedure	42 (11.4)	34		123
			Active trunk exercise program	41 (12.8)	41		112
Chok et al, Singapore	Physiotherapy + orthopaedic clinics + A/E	20-55	Physical therapy (endurance exercise at the PT department) + back hot pack	38 (9.7)	20	4 (2)	38
			Back hot pack (Home)	34 (8.1)	29	4 (2)	28
Constant et al, France	General practice	Unlimited	Spa therapy & usual GP care				63
			Waiting list group & usual GP care				63
George et al, USA	Physical therapy	18-55	Standard care physical therapy	37 (10.1)	53	4 (2.3)	32
			Fear-avoidance based physical therapy	40 (10)	62	4 (2.3)	34
Geisser et al, USA	University spinal programme	18-65	Manual therapy & Specific exercise (self corrections, stretching, strengthening)	39 (12.8)	67	284 (493)	26
			Sham Manual therapy & Specific exercise	39 (9.4)	56	370 (447.8)	25
			Manual Therapy & Non-specific exercise	37 (14.4)	80	370 (476.1)	24
			Sham Manual Therapy & non-specific exercise	46 (9.5)	61	284 (305.1)	25
Glomsrod et al, Norway	Physicians clinics and general population	18-50	Active back school (Lectures and back exercises)	41 (6.1)	65		37
			Usual medical care	39 (6.6)	57		35
Goldby et al, UK	General practice + hospital physicians	18-65	Spinal stabilisation & Attending the back school	43 (10.7)	68		84
			Manual therapy & Attending the back school	41 (11.7)	69.9		89
			Education (Booklet: Back in action) & Attending the back school	42 (13)	67.5		40
Hay et al, UK	General practice	18-64	A brief programme of pain management (general fitness and exercise at clinic and home, explanation about pain mechanisms, distress, encouragement of positive coping strategies, overcoming fear of "hurt=harm", and implementation of a graded return to usual activities)	40 (12.0)	50		201
			Physiotherapy including manual therapy techniques	41 (11.6)	55		201

Heymans et al, Netherlands	Occupational healthcare	18-65	Usual Dutch occupational physician care	41 (9.6)	17	35	103
			Low intensity back school	41 (10.2)	22	35	98
			High intensity back school	40 (9.5)	23	35	98
Hseih et al, USA	General population	18 +	Joint manipulation & myofascial therapy	48 (13.7)	33	12 (7.2)	52
			Joint manipulation	47 (14)	33	12 (7.2)	48
			Myofascial therapy	49 (14.8)	33	12 (6.8)	51
			Back school	48 (13.7)	40	11 (6.6)	48
Hurley et al, UK	Physiotherapy + General practice + self referral	18-65	Manipulation therapy (Passively move intervertebral joint within or beyond its range)	40 (11.6)	57	7.5 (3.1)	80
			Interferential therapy (Electrical stimulation)	40 (12.1)	62	7.6 (3)	80
			Manipulation & interferential therapy	41 (11.3)	60	8.3 (2.8)	80
Hurwitz et al , USA	Managed care facility	18+	Chiropractic care only	52 (16.5)	49		169
			Chiropractic care & physical modalities (Heat/cold, USS)	54 (16.8)	58		172
			Medical care (excluding physical treatment) only	49 (16.5)	47		170
			Medical care & physical modalities (Heat/cold, USS)	49 (16.7)	54		170
Hurwitz et al, USA	Network of healthcare	18 +	Chiropractic care only	52 (16.5)	49		340
			Chiropractic care & physical modalities (Heat/cold, USS)	53 (16.8)	58		340
Jellema et al, Netherlands	General practice	18-65	Minimal intervention strategy (Assessing psychosocial risks, providing information on back pain and treatments & advice on self care)	43 (11.1)	48	1.7	143
Kaapa et al, Finland	Occupational healthcare	22-75	Multidisciplinary rehabilitation: guided, group programme. : CBT, relaxation, back school education & physical therapy	46 (7.9)	100	72 (112.5)	59
			Individual physiotherapy	47 (7)	100	63 (103.5)	61
Kankaanpaa et al, Finland	Occupational healthcare		Active rehabilitation: guided exercises in a dept + behavioural support	40 (8.5)	34		30
			Passive treatment: which they considered as minor to the active arm, e.g. massage and thermal treatment	39 (7.3)	33		24
Karjalainen et al, Finland	General practice	25-60	Mini-intervention (Specific back exercises, reduce patient concerns & encourage physical activity)	44	59		56
			Mini-intervention & worksite visit	44	57		51
			Usual GP care	43	60		57
			Usual GP care				
Kerr et al, UK	General practice		Acupuncture	43 (11.5)	50	86 (84.9)	30
			Placebo TENS (non-functioning)	43 (12)	65	73 (77.4)	30
Kuukkanen et al, Finland	Occupational healthcare		Intensive training: intensive progressive exercises guided at the gym + home exercises		62		29
			Home exercise only: same as intensive, but unguided		48		29
			Control: usual activities, no trial exercises		54		28

Leclaire et al, Canada	Private physiatrist clinic	18-50	Standard care (rest, analgesics, physio) & Swedish back school	32 (7.7)	43		82
			Standard care (rest, analgesics, physio)	32 (8)	41		86
Lindstrom et al, Sweden	Occupational healthcare		Swedish back school & workplace visit + graded exercise (CBT approach)		24		51
			Usual care: rest& analgesics & physical treatment		38		52
Linton et al, Sweden	Mixed: General practice + General population	18-60	Back pain pamphlet	45	71		70
			Comprehensive information package	44	74		66
			CBT intervention	44	70		107
Mannion et al, Finland	General population	< 65	Modern active individual physiotherapy: strengthening, coordination and aerobics exercises, instructions on ergonomic principles + home exercises	46 (10.1)	61	520 (468)	46
			Muscle reconditioning on training devices (small groups)	45 (9.7)	54	504 (473.2)	47
			Low impact aerobic/stretching (groups)	44 (10.1)	55	676 (520)	44
Maul et al, Switzerland	Occupational healthcare	20-55	Back school & exercise	38 (8)			97
			Back school	39 (10)			86
Mehling et al, USA	General practice	20-70	Breath therapy	50 (12.1)	70	51 (26.6)	16
			Physical therapy: soft tissue mobilisation, joint mobilisation and exercises	49 (12.5)	58	57 (26.6)	12
Moseley et al, Australia	Mixed: Physiotherapy + General practice		Physiotherapy	43 (7)	64		29
			Usual GP care	38 (7)	54		28
Niemisto et al, Finland	General population	24-46	Manipulation, exercise & physician consultation	37 (5.6)	55	312	102
			Physician consultation only	37 (5.6)	53	312	102
Nordeman et al, Sweden	Mixed: General practice + physical therapy dept		Early access to physio (Individualised, exercise, advice, group education)	39 (12.1)	63		32
			Waiting list control	41 (11.1)	50		28
Peloso et al, USA	Outpatients	18 +	Tramadol & Acetamenophen combination tablets 375/325 2 PRN	58 (11.5)	64		167
			Placebo tablets 2 PRN	58 (13.6)	61		169
Rasmussen et al, Sweden	Physiotherapy	18-60	Stabilizing training (Individual) (Cognitive + stabilisation of spinal muscles)	39 (12)	70		24
			Manual treatment (Individual) (Other muscles exercises, no manipulation)	37 (10)	78		23
Rittweger et al, Germany	General population	40-60	Isodynamic lumbar extension	50 (6.6)	44	603 (520)	30
			Vibration exercise (On a machine with a vibrating platform)	54 (3.4)	52	754 (530.4)	30
Ritvanen et al, Finland	General population	20-60	Traditional chiropractic bone setting	41 (4.9)	45		33
			Physical therapy	42 (6.8)	43		28
Rossignol et al, Canada	Workers compensation board		Coordination of primary healthcare program	37 (9.7)	33		54
			Usual GP care	38 (10.2)	23		56
Ruoff et al, USA	Not clear	25-75	Tramadol/ Acetamenophen combination tablets 37.5/325 mg, 2 tablets	54 (11.9)	67		161
			Placebo tablets 2 tablets	54 (12)	59		157

Soukup et al, Norway	Mixed: General practice + General population + Referrals	18-50	Mensediesk exercise group intervention	40	53	676 (474.8)	34
			Waiting list group	40	49	578 (320.7)	35
Staal et al, Netherlands	Occupational healthcare		Graded activity (Physiotherapy + OT)	39 (9)	5	8.5	67
			Usual OT care	37 (8)	8	8	67
Torstensen et al, Norway	Social security offices	20-65	Medical exercise therapy (MET)	42 (11.2)	52		71
			Conventional physiotherapy (CP)	43 (12)	48		67
			Self exercise	40 (11.4)	51		70
Tsui et al, Hong Kong	Physiotherapy	20-55	Electro-acupuncture & back exercise	40 (12.1)	76	39 (31.8)	14
			Electrical heat acupuncture + back exercise	39 (9.2)	71	54 (54.5)	14
			Back exercise only	41 (8.3)	62	50 (41.6)	14
Turner et al, USA	General practice + Physicians + general population	20-65	Relaxation training (group)				24
			Cognitive therapy (group)				23
			Cognitive therapy & Relaxation training (group)				25
			Waiting list				30
Wand et al, UK	General practice + A/E patients		Assess & Advice & Physiotherapy	34 (9)	44		43
			Assess & Advice & wait	35 (7.9)	55		51
Werners et al, Germany	General practice		Interferential therapy: electrotherapy, to stimulate muscles fibres	38 (9.4)	43		68
			Motorised lumbar traction & massage	39 (9.5)	49		72
Yelland et al, Australia	General practice	21-70	Glucose lignocaine injection	52 (10.6)	59	769.6 (556.8)	28
			Exercise (Alternating: flexion and extension of spine and hips)	49 (10.4)	55	718 (483.6)	26
			Saline injection	50 (9.8)	56	759 (499.2)	27
			Normal activity	51 (11.2)	58	733 (551.2)	29

Appendix 7.1 Within-arm change in RMDQ, mean(SD), in trials' IPD

		Baseline		4weeks			6weeks			13weeks			27weeks			52weeks		
Trials	**Treatment arms**	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**Change Mean(SD)**	**n**	**Mean (SD)**	**Change Mean(SD)**	**n**	**Mean (SD)**	**Change Mean(SD)**	**n**	**Mean (SD)**	**Change Mean(SD)**	**n**	**Mean (SD)**	**Change Mean(SD)**
Hay et al	Manual physiotherapy	201	13.29 (4.88)							162	5.14 (5.83)	8.06 (6.03)				165	4.41 (5.49)	8.79 (6.09)
	Brief pain management	201	13.77 (4.82)							157	5.97 (5.93)	7.83 (6.63)				164	5.17 (5.75)	8.8 (6.37)
Heymans et al	Occupational health care	102	9.75 (5.04)							78	9.40 (6.55)	0.52 (10.22)	71	7.10 (6.27)	2.77 (9.53)	84	6.61 (5.91)	3.27 (9.26)
	Low intensity back school	94	7.93 (3.94)							73	8.41 (5.80)	0-.53 (8.20)	71	6.45 (6.09)	1.75 (8.53)	87	7.01 (6.54)	0.77 (8.73)
	High intensity back school	91	8.07 (3.89)							70	9.16 (6.46)	-0.82 (8.56)	66	7.94 (5.91)	0.40 (7.44)	73	6.56 (6.22)	1.60 (8.15)
Johnson et al	community-based treatment program of group sessions led physiotherapists	116	10.62 (3.91)							110	7.40 (5.33)	3.15 (5.64)	105	6.46 (4.67)	3.96 (4.45)	101	6.69 (5.64)	3.81 (5.44)
	Back pain educational booklet and usual GP care.	118	10.89 (3.96)							113	8.02 (5.28)	2.79 (4.63)	98	8.03 (5.35)	2.73 (4.77)	94	7.97 (5.48)	2.66 (4.80)
Moffett et al	Exercise programme	89	6.65 (4.01)				85	3.44 (3.34)	3.21 (3.57)				77	3.02 (3.59)	3.48 (4.66)	83	2.93 (3.42)	3.64 (4.37)
	Routine GP care	98	5.56 (3.94)				94	3.68 (4.00)	1.63 (3.78)				86	3.90 (4.43)	1.20 (4.56)	88	3.74 (4.48)	1.34 (4.24)
UKBEAM	Best GP care	338	8.97 (3.87)	268	7.52 (4.83)	1.29 (3.88)				256	6.58 (5.04)	2.10 (4.37)				248	5.92 (5.05)	2.80 (4.64)
	Exercise	310	9.21 (4.35)	239	6.72 (4.92)	2.26 (3.93)				225	5.42 (4.83)	3.56 (4.13)				216	5.99 (5.63)	3.23 (4.30)
	Manipulation	353	8.91 (3.98)	325	6.70 (4.98)	2.17 (4.08)				287	5.09 (5.03)	3.64 (4.78)				273	5.08 (5.17)	3.85 (5.03)
	Combined exercise and manipulation	332	8.99 (3.92)	286	6.50 (4.48)	2.41 (3.65)				259	4.84 (4.85)	4.07 (4.27)				258	4.77 (5.02)	4.07 (4.57)

Appendix 7.2 Within-arm differences in change in ODI scores (mean, 95% CI) according to the selected characteristics in Thomas et al trial.

			13weeks			52weeks	
Characteristics	**Treatment arms**	**n**	**Unadjusted**	**Adjusted**		**Unadjusted**	**Adjusted**
Age, *younger than 43 years*	**Acupuncture**	75:71	-4.30 (-8.97,0.38)	-4.07 (-8.86, 0.73)	69:65	-4.23 (-9.71, 1.25)	-4.09 (-9.76, 1.58)
	Usual care	34:37	-3.06 (-10.32, 4.20)	-2.92 (-10.37, 4.52	26:31	-0.08 (-7.82, 7.98)	-0.28 (-8.48, 7.93)
Gender, *male*	**Acupuncture**	53:93	-2.93 (-7.82, 1.97)	-2.36 (-7.35, 2.63)	49:85	-1.61 (-7.34, 4.12)	0.73 (-6.62, 5.17)
	Usual care	29:42	-2.83 (-10.21, 4.56)	-2.71 (-10.25, 4.91)	19:38	2.76 (-5.55, 11.08)	2.94 (-5.71, 11.59)
History of LBP, *absent*	**Acupuncture**	22:124	1.80 (-4.80, 8.40)	2.77 (-3.85, 9.38)	22:112	-0.64 (-8.11, 6.82)	0.07 (-7.77, 7.61)
	Usual care	12:59	-0.84 (-10.57, 8.88)	-1.87 (-11.82, 8.09)	10:47	0.79 (-9.55, 11.14)	1.22 (-9.46, 11.89)

n: number in the comparable groups.

Appendix 8 Literature search strategy for cohort studies

Back pain_Prognosis_Cohorts_Primary care (12/04/2011)

1. exp Low back pain/ OR exp back pain/
2. exp Pain/
3. (back AND pain).ti,ab [Limit to: Core clinical journals (AIM) and Humans and English Language]
4. 1 AND 2 AND 3

5. exp Prognosis/
6. exp disease progression/
7. predict.ti,ab
8. factor.ti,ab
9. model.ti,ab
10. evolution.ti,ab
11. history.ti,ab
12. course.ti,ab
13. determinant.ti,ab
14. pattern.ti,ab
15. screen.ti,ab
16. long-term.ti.ab
17. progress.ti,ab
18. modif.ti,ab
19. mediate.ti,ab
20. OR/ 4-16
21. 4 AND 20

22. (epidemiology AND studies).ti,ab
23. exp Epidemiology studies/
24. cohort.ti,ab
25. retrospective.ti,ab
26. prospective.ti,ab
27. longitudinal.ti,ab
28. inception.ti,ab
29. observation.ti,ab
30. outcome.ti,ab
31. OR/ 22-30
32. 21 AND 31

33. exp Primary Health Care/
34. exp Family Practice/
35. exp Physicians, Family/
36. exp Community Health Services/
37. "General Practice".ti,ab
38. "family practice".ti,ab
39. "family physician".ti,ab
40. AMBULATORY CARE/
41. “ambulatory adj2 care”.ti,ab
42. OR/ 33-41

43. 32 AND 42