Chapter 1: Literature Review

A systematic search of the electronic databases of MEDLINE, EMBASE, CINAHL and the Cochrane Central Register was conducted using the following keywords: Musculoskeletal, Low Back Pain, Neck Pain and Prevalence in the U.K. There were only 2 articles for low back pain and one for neck pain which happened to be one of the same papers found for back pain. The search was opened to omit the U.K. Searches were done on the NHS website and that of the Department of Work and Pensions but very little current data exists for the prevalence of back and neck pain. Although there appeared to be 153 papers for a search on guidelines and back pain and 27 for guidelines and neck pain, once the abstracts had been read it was apparent that only 18 papers concerned the development and comparison of guidelines for back and neck pain. 'Subgrouping and prognostic studies' revealed 4 studies for back pain and none for neck pain but searching the references of those papers gave a total of 15 papers in this much talked about but underdeveloped aspect of musculoskeletal research.

Prognostic studies for back and neck pain has centred around a handful of researchers and 39 papers were identified which dealt with prognostic factors for back and neck pain although the search initially looked as though there were 230 papers. All papers that were written on prognostic factors were included in the literature review and their limitations were commented upon. A search for the Bournemouth Questionnaire revealed 163 papers but only 11 actually related to the development, validation and translation of the questionnaire. There were 14 papers developing the concept of clinical as opposed to statistical significance. Nine papers were identified for calculating anchor-based clinical significance and eleven papers have been written for determining the distribution-based

method. These papers compared the advantages and disadvantages of these two methods for determining clinically significant improvement.

1.1 Preamble

Musculoskeletal problems are one of the most common reasons adults seek medical care with between twenty to forty percent of adults consulting their general practitioner (GP) over the course of a year (Savingy, Kuntze, Watson, Underwood, Richie, Cotterell et al 2009; Keijsers, Feleus, Miedma, Koes, Bierma-Zeinstra 2010; Jordan et al 2010). The most common site of musculoskeletal pain is in the low back, followed by head/neck pain and upper and lower limb pain (Jordan et al 2010). The prevalence of back and neck pain in Britain varies between twenty-nine percent and forty-five percent of the population and this prevalence seems to increase with age (twenty-three percent of 18-24 year olds and fifty percent of 55-64 year olds) (Andersson 1999; Webb et al 2003; Parsons et al 2007; Savingy et al 2009; Keijsers et al 2010).

Successive studies have found the main reasons patients consult chiropractors is for low back pain, neck pain and headaches (Assendelft, Pfiefle & Bouter 1995; Mootz et al 2005; Ailliet, Rubinstein & de Vet 2010). Neck pain is often associated with shoulder pain and less often with upper limb pain (Eltayeb, Staal, Kennes, Lamberts & de Bie 2007; Feleus et al 2008; Keijsers et al 2010). Moreover, primary headaches of which migraine, tension and cervicogenic account for the majority, are often associated with neck pain and have been shown to respond to spinal manipulative therapy of the cervical spine (Haas, Spegman, Peterson, Aickin & Vavrek 2010). Low back pain is often associated with leg pain (sciatica) with a lifetime prevalence of forty percent and whilst a considerable body of evidence exists to support the use of spinal manipulative therapy for back pain, there is less evidence for

patients with leg pain although that which does exist suggests spinal manipulation is beneficial (Santilli, Beghi & Finucci 2006; Lawrence et al 2008; Murphy, Hurwitz & McGovern 2009; Schulz et al 2011).

Patients with less than six-weeks duration of back pain (acute) are advised to remain active, avoid bed rest and use over-the-counter pain medication, but the correct management of patients whose symptoms persists for more than six-weeks (subacute) or more than twelve-weeks (chronic) is less clear (Savingy et al 2009). Around forty percent of musculoskeletal patients have chronic pain which is often associated with work disability (Bergström, Hagberg, Bodin, Jensen & Bergström 2011; Bergström, Jensen, Hagberg & Bergström 2011). Little is known about the current economic burden of chronic back and neck pain in the United Kingdom despite the fact that it is repeatedly cited as the second most common cause of the 2.6 million people claiming employment and support allowance (formerly incapacity benefit) after that of mental health conditions. It is known that back pain disability rose by one hundred and four percent between 1986 and 1992 and that 116 million production days were lost related to back pain (low back, neck and thoracic) in 1994/5, having risen from 106 million days in 1992 (Hillman, Wright, Rajaratnam & Chamberlain 1996; Maniadakis & Gray 2000). The incidence of back, neck and thoracic pain does not necessarily increase with age but the duration of the episodes of pain does increase with age, and women are more likely to report pain than men (Leboeuf-Yde, Nielsen, Kyvik, Fejer & Hartvigsen 2009; Jordan et al 2010).

The total social cost of back pain to Britain in 1993 was £6 billion and it is estimated that the cost of low back pain is rising by £500 million annually (Hillman et al 1996). It would appear that the prevalence of low back pain is much higher than forty years ago (Harkness, Macfarlane, Silman & McBeth 2005). Ten million working days were lost in Britain because of

low back pain in 1955, twenty-five million days in 1985 and seventy-five million days in 1994 (Troup 1996). No current estimates of the costs of back pain in the UK exist in the literature but estimates of the direct health care costs of back pain in 1998 were £1362 per patient in the UK (Maniadakis et al 2000) but as the retail price index has increased by more than thirty percent since that time, it is likely that the published figures under estimate the current economic burden (Savingy et al 2009). In 1998 total health care costs for back pain were £1632 million of which £565 million was being spent by patients outside the NHS on osteopathy, physiotherapy, acupuncture and chiropractic, but the indirect costs of back pain due to lost production were much higher and estimated to be up to £9090 million (Meniadakis et al 2000) .

In the United States the cost of back pain is $100 billion a year, much of which is spent on prolonged courses of treatment for those patients who fail to improve and become disabled (Foster, Dziedzic, van der Windt, Fritz & Hay 2009). A prevalence of between six and eleven percent for chronic low back pain has been established with the cost of low back pain available for Germany, in excess of €7000 per patient with seventy five percent of the cost being work absenteeism (Juniper, Le and Mladsi 2009). However, a study of Scandinavian chronic back and neck pain patients referred for tertiary rehabilitative care has suggested a cost of rehabilitation together with the patient's lost productivity could amount to as much as €94 494 per patient (Jensen et al 2009) . In Sweden back pain is estimated to cost one point three percent of Gross Domestic Product (Bergström et al 2011a).

In Britain a Government Statistical Service survey did find fifteen percent of back pain sufferers had back pain throughout the year in 1998 and forty percent of them consulted their GP with ten percent visiting a chiropractor, osteopath or acupuncturist and thirteen percent of those patients had taken time off work because of their back pain (Great Britain,

Government Statistics Office 1999). Disability caused by low back pain affects approximately one-quarter of adults in any one year and is the most common physical disability in the working population of the UK (Maughan & Lewis 2010). There is a commonly held belief that ninety percent of patients with back pain recover within six weeks but this might be too optimistic as it is at odds with the findings of several authors who have found that although most back pain patients return to work within three months, pain and disability take much longer to resolve with twenty eight percent of patients' symptoms not resolved at one year (Hestbaek, Lebouef-Yde & Manniche 2003; Henschke et al 2008; Koes et al 2010; Bergström et al 2011a). Other authors have suggested that a large proportion of back pain sufferers still experience pain at 12 months (up to seventy five percent) and that nearly seventy eight percent of these patients experience relapses (Dionne et al 2008; Koes et al 2010, Kongsted & Lebouef-Yde 2010).

One of the reasons that there has been little agreement in the literature over the prevalence of back pain is that, until recently, there was no consensus in the literature over what constituted low back pain (Dionne et al 2008; Koes et al 2010). Through a modified Delphi study these authors sought to define location, duration and severity which would be the minimal classification of back pain (low back only, and bad enough to limit one's daily activities for at least one day). Moreover, thanks to the revised management of back and neck pain following the publication of the "Biopsychosocial Model of Back Pain (BPS model)" (Waddell 1987), adapting the psychiatrists' model of mental health to back pain, it would appear that there are encouraging signs that back and neck disability rates have shown signs of decreasing in recent years. In the Netherlands occupational back disability as a result of back pain decreased by thirty seven percent in men and twenty one percent in women between 1985 and 1999/2000 (Steenstra, Verbeek, Prinze & Knol 2006). In Sweden the

incidence of neck and shoulder pain with concurrent complications decreased between 2002 and 2006 with sixteen percent rather than nineteen percent of these patients being off work, although these authors concluded it was too early to conclude the 'epidemic' had peaked (Leijon, Wahlström & Mulder 2009). Notwithstanding this, back pain, neck pain and headaches remain significant in terms of individual and societal impact, and their management continues to tax clinicians and researchers alike. This has led to back and neck pain being subjected to systematic reviews and clinical guidelines to inform treatment approaches (Bolton & Breen 1999).

1.2 Development of Guidelines for Back Pain

The BPS model states the back pain experience is a function of interacting combinations of patho-anatomical, neuro-physiological, physical and psychosocial factors which are different for each patient (Maughan & Lewis 2010). Following Waddell's groundbreaking paper (1987) and the resulting report of the Clinical Advisory Standards Group Report (Rosen 1994), guidelines for the management of low back pain, using the available evidence, started to emerge around the world between 1994 and 2000. Difficulty has arisen when comparing these guidelines from thirteen countries and two international European guidelines as some were written for acute/subacute back pain of less than twelve-weeks duration (USA, New Zealand, Australia and UK) whilst others focussed on chronic back pain (Netherlands, Switzerland, Denmark, Sweden and Israel) (Koes, van Tulder, Ostello, Burton & Waddell 2001).

Researchers and clinicians accept that low back (and neck pain) are not simply acute or chronic but fluctuate over time with frequent recurrences or exacerbations (Enthoven, Skargren & Öberg 2004; Dionne et al 2008; Leijon et al 2009; Kongsted & Lebouef-Yde 2010).

A consensus has emerged regarding diagnosis, a history and examination to exclude serious pathology or neurological deficit, consideration of psychosocial complications, and the fact that radiographs are not useful for nonspecific low back pain. Patients have been reassured, advised to remain active, to progressively increase activity levels, to use analgesia or non-steroidal anti-inflammatory drugs where recommended, to avoid bed rest, to consider spinal manipulation in the short-term and tailored exercise therapy (Hildebrandt et al 2004).

The guidelines, based on the BPS model, treating back pain as an illness rather than a disease, produced by the Royal College of General Practitioners (Waddell, McIntosh, Hutchinson, Feder & Lewis 1999) have been superseded by the European Guidelines (2004) (Hildebrandt et al 2004; Van Tulder et al 2004), The Musculoskeletal Services Framework (Department of Health 2006) and more recently, the National Institute for Clinical Excellence (NICE) guidelines for the management of low back pain (Savingy et al 2009). Consideration of spinal manipulation, acupuncture (dry needling) and tailored, supervised exercise programmes in the management of low back pain have remained in the guidelines. A recent comparison of non-specific low back pain guidelines published between 2000 and 2008 has shown little has changed in recommendations for the management of back pain in the last decade since these authors published their previous review of the guidelines published up to and including the year 2000, although their implementation remains a challenge in both clinical practice and research (Koes, Van Tulder, Ostelo, Burton & Waddell 2001; Koes et al 2010). Researchers look at the data of clinical trials to draw up guidelines but clinicians, although they tend to agree with the guidelines for back pain, have to contend with patient preferences, lack of access to the multimodal approach proposed by the guidelines and their clinical judgement based on their knowledge of the patient (Schers, Wensing, Huijsmans, van

Tulder & Grol 2001; Verbeek, Sengers, Riemans & Haafkens 2004; Chenot et al 2008; Phelan, van Ryn, Wall & Burgess 2009).

A limitation of guidelines for musculoskeletal complaints is that they tend to be "one size fits all" in their approach (Foster et al 2009). Adherence to guidelines has been hampered by patient expectation and prior experience, the clinical experience of the physician and the fact that doctors are inclined to give in to patients' demands irrespective of guideline recommendations (Schers, Braspenning, Drijver, Wensing & Grol 2000; Schers et al 2001; Chenot et al 2008). At the time of development of back pain guidelines in Britain some GPs were given limited purchasing power under the Fund-Holding Scheme in the 1990s. Local arrangements were made between chiropractors and local GP practices whereby some patients were able to receive spinal manipulation. The change of government in 1997 saw the demise of this scheme which was replaced by a collective buying scheme through Primary Care Trusts which resulted in very little availability of chiropractic treatment to patients through the NHS. Barriers to routine referral of back pain patients to chiropractors centre around a lack of formalised referral relationships and medical practitioner concerns about efficiency, continuity, quality and patient safety in healthcare (Greene, Smith, Allareddy & Haas 2006). Many of these concerns have been allayed when chiropractors have been introduced into a healthcare team (Garner et al 2008). Despite the majority of the chiropractic profession being in favour of some sort of future partnership with the NHS, co-operation is sporadic and limited by the lack of an overall policy plan, requirement for firm evidence of efficacy and constraints placed on NHS purchasers (Leboeuf-Yde, Andrén, Gernardt & Malmqvist 1997; Langworthy, Breen, Vogel & Collier 2002).

1.3 Subgrouping patients

Despite many years of guideline development, back and neck pain remain a significant problem in terms of healthcare and disability which raises the question of whether guidelines for the management of back and neck pain work better for some groups of patients than others? Back, neck and shoulder pain is usually managed in primary care by general practitioners, physiotherapists, chiropractors and osteopaths (Henschke et al 2008; Foster et al 2009; Foster, Hill & Hay 2011). Presently there are no easy answers on how best to treat these conditions as it seems a number of different treatments have an effect, albeit to a limited degree (Kongsted & Lebouef-Yde 2010). Most trials testing non-pharmalogical interventions, such as chiropractic (spinal manipulative therapy), for these conditions have shown small to moderate benefits, a finding which is often at odds with the experience of practitioners who see patients improve, often dramatically (Bolton 2003; Wyatt, Underwood, Scheel, Cassidy & Nagel 2004; Henschke et al 2008; Thiel & Bolton 2008; Foster et al 2009, Leboeuf-Yde et al 2009; Foster et al 2011; Van Middelkoop et al 2011).

It has been stated that large, pragmatic trials with broad eligibility criteria and high inclusion rates provide the most reliable data on the effects of treatment. However, these trials often fail to demonstrate superiority of one conservative treatment over another for musculoskeletal complaints (Peat 2008; Hayden et al 2010). Successive trials, performed on these heterogeneous groups of patients could be the reason as it is unlikely that one intervention will be effective for all patients with diverse demographic and clinical characteristics. These interventions have been shown to be more effective when directed at selected (homogeneous) groups of patients but there is a risk of introducing bias when the study population is not heterogeneous (Hancock, Herbert & Maher 2009). In an attempt to break this stalemate, a number of researchers have showing an interest in the study of

9

subgroups of back pain sufferers with preliminary results suggesting interventions directed at subgroups of patients are more effective than interventions directed at mixed populations of low back pain patients (Enthoven, Skargren & Öberg 2004; Hall, McIntosh & Boyle 2009).

Recommendations for the use of spinal manipulative therapy and mobilisation have been made for low back and neck pain in current guidelines although there are not many high quality studies that differentiate acute and chronic patients. Many of these studies upon which the guidelines are based only have short-term follow-up periods which fail to capture the fluctuating nature of the back/neck pain experience. It has been suggested that future studies look at well-defined subgroups of patients, establish optimal number of treatment sessions, cost effectiveness and adverse side-effects of care (Brontfort, Haas, Evans & Bouter 2004; Peat 2008; Bergström, Bergström, Hagberg, Bodin & Jensen 2010; Brontfort, Haas, Evans, Leininger & Triano 2010). Increasingly researchers and clinicians are becoming convinced that effectiveness of interventions can be improved by targeting the provision of specific interventions at patients who respond best to that treatment (Wernecke & Hart 2004; Fritz, Childs & Flynn 2005; Steenstra, Verbeek, Prinze & Knol 2006; Hancock et al 2009; Pransky, Borkan, Young & Cherkin 2011). In effect, it has been suggested that there are specific types of 'treatment responders' within the broad domain of non-specific back and neck pain patients.

Clinical prediction rules were originally used to quantify the usefulness of clusters of patient characteristics (for example, history and examination findings) for diagnosis and prognosis (Hancock et al 2009). A growing body of evidence has emerged concerning the development of clinical prediction rules as a means of identifying patients according to likely outcome and responsiveness to treatment, in other words, patients who will respond best to a certain intervention (Peat 2008, Hancock et al 2009). There is a need for these prediction rules to be

implemented carefully as methods used to identify subgroups could lead to biased and misleading results (Hancock et al 2009; Foster et al 2011).

As there is complexity of the interactions between the patho-anatomical, neuro-physiological and the psychosocial aspects of the back and neck pain experience for each patient, the effects of treatment intervention may well be diluted in trials of heterogeneous groups of patients who probably have diverse treatment needs (Maughan & Lewis 2010). An example of the complexity of these interactions is how the UK BEAM trial revealed that age, work status, age at leaving school, pain and disability, quality of life and beliefs were prognostic factors for recovery from back pain but these factors did not predict response to spinal manipulation, exercise or a combination of both interventions in this heterogeneous study population (Underwood, Morton & Farrin 2007).

Moreover, whilst we have amassed evidence on the 'bio' and 'pyscho' aspects of the pain experience, apart from work-related factors, the 'social' aspect has been somewhat neglected with little or no information on factors such as societal benefits, social support, benefit systems, family influences and the availability of appropriate healthcare. These might be fruitful avenues of investigation. However, a study of sixty-nine low back prediction studies, looking at 221 distinct predictors, was made and these authors found that social factors are often seen as out of the control of the patient or researcher and have been put in the 'too-difficult pile' (Hayden et al 2010).

Characteristics that identify subgroups of patients who respond differently to treatment are known as 'treatment effect modifiers' whereas 'prognostic factors' are characteristics that identify patients who recover at different rates or have different outcomes irrespective of treatment (Hancock et al 2009). Generally prognostic factors sometimes overlap with

treatment effect modifiers but this is not always the case (Foster et al 2011). Clinical Prediction Rules quantify clusters of patient characteristics that assist diagnosis and prognosis for different treatments (Peat 2008). This is a three stage process: derivation, where studies have been conducted to determine variables that predict outcome, validation, where these variables have been tested on new patients in a new setting, and impact analysis, where one investigates whether or not clinical prediction rules in clinical practice improve patient outcomes (Hancock et al 2009; Moons, Royston, Vergouwe, Grobbee & Altman 2009; Hayden et al 2010).

If one is trying to identify factors associated with treatment outcome, the design of the study must be experimental with a control group that does not undergo any intervention (Steenstra et al 2006; Hancock et al 2009; Kongsted & Leboeuf-Yde 2010). Single-arm trials cannot provide estimates of treatment effect (potential treatment modifiers) although single arm studies can indicate potential prognostic factors (particular clinical presentation associated with good outcome regardless of treatment) by tracking patients over time (Flynn et al 2002; Steenstra et al 2006; Carroll et al 2009; Foster et al 2009; Hancock et al 2009; Foster et al 2011). A study of sick-leave related to back pain found patients who had a perceived reduced ability to work and a belief that work would aggravate their back pain were prognostic factors for sick-leave but perceived reduced ability to work, constant back ache when working and a high score for gastrointestinal complaints acted as treatment effect modifiers demonstrating prognostic factors and treatment effect modifiers can be different but not mutually exclusive (Hagen, Svensen & Eriksen 2005).

Clinicians often believe they can identify subgroups of patients that will respond differently to an intervention for musculoskeletal pain. There is early evidence that it might be possible to better match the patient with treatment intervention for low back pain and calls for sub-

grouping to find out what treatment works best and for whom have been around since 1998 (Borkan, Koes, Reis & Cherkin 1998; Foster et al 2011). GPs are better at predicting patients who will have an unfavourable outcome than those who will improve, suggesting primary care practitioners may be well-placed to identify those patients who need referral for more extensive treatment early in the course of an episode of low back pain (Hayden et al 2010). However, there is little consensus in the literature about how one sets about identifying subgroups of back pain patients and few methods have been subjected to rigorous testing. Causal homogeneity does not imply prognostic homogeneity or treatment responsiveness. Furthermore, none of the guidelines for the management of back pain recommend subgrouping patients (Foster et al 2011). Nevertheless, research to identify subsets of low back pain patients that respond more favourably to chiropractic needs to be conducted so that treatment can directed at patients who will benefit most to improve both patient outcomes and cost-effectiveness for the individual or third party payers (Breen 2010).

A multicentre prospective observational study, exploring the six-month clinical course with measurements of 'bothersomeness' using hierarchical cluster analysis, has resulted in the first attempt within the chiropractic profession to divide patients into four clusters. These were the stable (course relatively unchanged over time), fast improvers, typical patients (medium bothersomeness at baseline and average improvers) and slow improvers. If future studies can validate these clusters in larger cohorts, it has been suggested the most appropriate treatment strategies for each group should be investigated to ensure optimal efficacy of treatment for non-specific low back pain (Axén et al 2011).

Other researchers have recently attempted to classify patients at risk of developing chronic, disabling neck and low back pain. Based on psychological and psychosocial characteristics, these authors divided patients into adaptive copers, interpersonally distressed (low levels of

social support) and dysfunctional (high pain severity, high disability and high levels of affective distress) in this prospective cohort study over thirty-six months. Multidisciplinary approaches were found to help the dysfunctional and adaptive copers whereas the interpersonally distressed patients responded least well and had most sick-leave (Bergström et al 2011a; Bergström et al 2011b). Patients in these studies were offered physical treatment and Cognitive Behavioural Therapy, addressing physical and psychosocial aspects of the pain experience and as such management of the interpersonally distressed individuals remains elusive. This study had its limitations in that there were no data concerning pain severity or duration which together with the work-related factors would have enhanced the internal validity of the study.

1.4 Prognostic Factors

Prognostic factors are the characteristics that identify patients who will recover at different rates or will have different outcomes, regardless of treatment (Carroll et al 2009). Baseline prognostic factors can be divided into general, health and work-related characteristics for the individual (Hayden et al 2010). Prognostic factors for neck and back pain are either non-modifiable (e.g. age, gender) or potentially modifiable (e.g. coping strategies, levels of exercise). The challenge is how to disentangle these physical and psychosocial characteristics and the interactions between them; to relate physical, demographic and psychosocial characteristics to find subgroups of patients who might respond to different interventions, improving their prognosis (Waddell 2005). Until 2009 it would appear no one had classified predictors of outcome for back pain. Hayden et al (2009) used five categories: (i) Characteristics of the current episode of back pain, (ii) Individual characteristics, (iii) Psychological characteristics, (iv) Work environment and (v) Social environment. For the

purpose of this review the alternative classification system proposed by Carroll et al (2009) has been used:

1. Demographic and socioeconomic: either non-modifiable (e.g. age, gender) or not easily modifiable (socio-economic status)

2. Health factors and pain history: function and disability in a BPS approach, taking into account environmental (society, attitudes and legislation) and personal factors which are not easily modifiable by the individual in this context (Dahl 2002).

3. Workplace factors: Job tasks and ergonomics of the workplace-potentially modifiable.

4. Psychological and social factors: e.g. anxiety, depression, coping strategies and interpersonal factors (relationships with friends)-potentially modifiable

5. Societal factors: Compensation systems and laws-potentially modifiable but not by the individual.

6. Genetic factors: Not modifiable.

7. Health behaviours: e.g. levels of exercise-potentially modifiable.

1. **Demographic and Socioeconomic Factors (non-modifiable)**

    For back or neck pain patients, demographic features such as age has often been cited as a predictor of outcome and authors usually state the older the patient, the less favourable the outcome is likely to be (Leboeuf-Yde et al 2004; Michaelson, Sjölander & Johansson 2004; Hill, Lewis, Sim, Hay & Dziedzic 2007; Mallen, Peat, Thomas, Dunn & Croft 2007; Henschke et al 2008; Schellingerhout et al 2008). It has been found that being female is predictive of disability in the long-term (Leboeuf-Yde et al 2004; Enthoven et al 2006; Hansson, Hansson & Jonsson 2006; Carroll et al 2009;

Kongsted & Leboeuf-Yde 2010; Axén et al 2011). Whilst some authors cite being female affects prognosis, these findings are at odds with the findings of other researchers who have found no link with gender and prognosis (Denison, Åsenlöf & Lindberg 2004; Michaelson et al 2004). Higher levels of education have been cited on occasion as having a helpful effect on prognosis of back and neck pain sufferers (Koleck, Mazaux, Rascle & Bruchon-Schweitzer 2006; Rubinstein et al 2008) and being a blue collar worker had the opposite effect on prognosis for both back and neck pain patients (Karjalainen et al 2004; Hill et al 2007; Carroll et al 2009; Bergström et al 2011a).

Evidence that a link exists between smoking and low back pain has been cited in the literature for over a decade but a causal link has not been established. It is likely that the duration and frequency of back pain is increased by being a smoker and that abstinence could be a useful means of preventing certain kinds of back pain (Leboeuf-Yde, Yashin & Lauritzen 1996; Leboeuf-Yde, Kyvik & Bruun 1998; Zvolensky, McMillan, Gonzalez & Asmundson 2009). At best, smoking can only be regarded as a weak risk factor for developing low back pain although it offers no information about prognosis (Leboeuf-Yde 1999). Similarly, although much interest in lifestyle factors affecting back pain has been made, no clear link between alcohol consumption and low back pain has been made (Leboeuf-Yde 2000; Ndetan, Bae, Evans, Rupert & Singh 2009; Holmberg and Thelin 2010). Some authors have found that moderate alcohol consumption reduces the likelihood of sick leave as a result of back or neck pain but their findings were confined to female public sector workers (Skillgate, Vingård, Josephson, Holm & Alfredsson 2009). An increased Body Mass Index, obesity and

living in an area of social deprivation has also been linked with intense, chronic neck and back pain (Webb et al 2003).

2. **Health Factors and Pain history (non-modifiable)**

Pain intensity has been frequently cited as having an effect on the prognosis for back and neck pain sufferers, particularly in the short-term (Mallen et al 2007; Schellingerhout et al 2008; Ssavedra-Hernádez et al 2011). These authors have found neck pain patients with accompanying back pain had a less favourable long-term prognosis. Similar findings had been reported in a previous study (Hill et al 2007). Moreover, it has been found that accompanying neck pain affected the prognosis in back pain patients (Leboeuf-Yde et al 2004; Leboeuf-Yde et al 2005; Enthoven et al 2006).

Patients with back pain who respond to chiropractic treatment normally do so very quickly by the fourth visit with very little further improvement at three and twelve months. However, these improved patients do experience recurrent episodes (Axén, Rosenbaum, Röbech, Wren & Leboeuf-Yde 2002, Leboeuf-Yde et al 2004; Leboeuf-Yde et al 2005; Enthoven et al 2006). Thiel & Bolton (2008) found that patients who presented with neck pain and/or shoulder/arm pain, reduced neck, shoulder/arm movement, headache, upper/midback pain often had immediate improvement after cervical manipulation. Patients with the additional symptoms of numbness/tingling in the upper limb, fainting/dizziness or light-headedness, often associated with more chronic syndromes, were more likely to feel worse immediately after cervical manipulation (Michaelson et al 2004, Rubinstein et al 2008; Thiel & Bolton 2008). Patients undergoing chiropractic treatment for neck and back pain had a more

favourable prognosis if they had a short duration of pain prior to presentation (Fritz, et al 2005; Rubinstein et al 2008).

A previous systematic review by Borge, Leboeuf-Yde & Lothe (2001) found no link between physical examination findings and treatment outcome in back pain. A recent study has found that restricted cervical extension and negative upper limb tension tests affected prognosis but these findings had to be present with other demographic and health factors which calls into question the validity of physical parameters as independent variables for assessing prognosis in musculoskeletal conditions (Ssavedra-Hernádez et al 2011). Similarly, patients with signs of disc problems on examination and the presence of leg pain had a worse prognosis than back pain patients of muscular origin although demographic and clinical characteristics such as patient gender, pain duration and severity had to be taken into consideration and it was these characteristics that accounted for the worsened prognosis rather than leg pain itself (Kongsted & Leboeuf-Yde 2010; Schulz et al 2011). Despite the presence of broad-based disc herniation and nerve root compromise on Magnetic Resonance Imaging (MRI), prognosis at fourteen months in male patients was good, despite follow-up MRI not demonstrating change (Jensen, Albert, Sorensen, Manniche & Leboeuf-Yde 2007).

It is at least twenty years since physical parameters were used as the primary outcome in back pain research. Bendix, Bendix and Haestrup (1998) were able to demonstrate that the endurance of the stabilising muscles of the back was predictive of recovery for patients in tertiary rehabilitation programs but subsequent research has failed to demonstrate physical parameters such as back endurance as predictors (Michaelson et al 2004). Patient experiences, attitudes and beliefs serve as a better

measure of whether or not a patient is improving, confirming long-held beliefs that physical parameters do not act as good measures of outcome (Bolton 1994; Deyo et al 1994; Michaelson et al 2004; Hamberg-van Reenen, Ariëns, Blatter, van Mechelen & Bongers, 2007; Hill et al 2007; Hill et al 2011). Increased physical functioning after an episode of back pain has been found to have more to do with a decrease in fear-avoidance beliefs and psychological distress than physical fitness (Bousema, Verbunt, Seelen, Vlaeyen & Knottnerus 2007).

For back pain patients, higher levels of pain intensity, longer duration of symptoms and perceived risk of persistence have been associated with longer recovery times (Hansson et al 2006; Hill et al 2007; Langworthy & Breen 2007; Mallen et al 2007; Henschke et al 2008; Keeley et al 2008; Jensen et al 2009). Furthermore, these authors found that the longer patients had been symptomatic or had been off work prior to treatment had an adverse effect on prognosis. Interestingly, other authors have found that sick leave and duration of symptoms did not have predictive value in assessing pain reduction in chronic neck and back pain sufferers (Denison et al 2004; Michaelson et al 2004; Kongsted & Leboeuf-Yde 2010). This is probably because dividing patients into acute and chronic might be too simplistic in some instances and using additional prognostic factors such as psychosocial domains, better reflecting the BPS model, may result in more acceptable labels for patients (Hayden et al 2010; Kongsted & Leboeuf-Yde 2010).

Rubinstein et al (2008) also found that patients not being on sick leave, not being tired, having a favourable treatment outcome expectation, not having morning pain and having a good perception of their general health were more likely to have a favourable treatment outcome. These findings support the work of previous authors

with regard to patient expectations of treatment outcome with regard to low back pain and neck pain showing that patients' general expectation of improvement was associated with improved functional status (Hall et al 2007; Myers et al 2007). Good general health status and a shorter total duration of pain in the previous year had also been found to have an effect on treatment outcome (Langworthy & Breen 2007; Leboeuf-Yde et al 2009). A history of previous episode/s, pain severity and longer periods of sick leave were all associated with a less favourable outcome in neck pain patients (Eltayeb, Staal, Hassan & de Bie 2007; Hill et al 2007; Carroll et al 2009; Bergström et al 2011a).

Some of the predictors identified in numerous research studies are at odds with the findings of other researchers. For example, Michaelson et al (2004) reported that variables such as history of sick leave and pain duration had little predictive value for chronic back and neck pain patients. Ozegovic, Carroll & Cassidy (2009) found pain intensity and health status prior to injury were associated with a more favourable outcome for whiplash patients. This finding was supported by Lindell, Johansson, Strender (2010) for non-specific spinal pain. Bot et al (2005) reported that symptom duration prior to presentation and a prior history of neck and shoulder pain affected outcome. Differences in predictors of outcome have probably been found because some studies are for acute patients, some chronic patients and follow-up periods vary. Interestingly one study found patients who had pain and disability at one year follow-up were often still consulting healthcare providers for episodes of back and neck pain at five year follow-up (Enthoven et al 2004).

3. **Workplace Factors (potentially modifiable)**

Low back pain of sudden onset at work has been strongly associated with exposure to specific patient-handling tasks/work-related accidents and has been associated with work absenteeism and short-term disability (Waddell 2005, Koleck et al 2006). Job satisfaction has been cited in the literature as a predictor of a favourable outcome from an episode of back pain (Dunn & Croft 2004). Neck and back pain patients with high quantitative demands at work, repetitive work, poor social support at work or poor work station design and sedentary work have been associated with a less favourable outcome (Hagen et al 2005; Eltayeb et al 2007; Côté et al 2008; Lakke, Soer, Takken & Reneman 2009; Holtermann, Hansen, Burr & Søgaard 2010, Bergström et al 2011a). Fear of movement was found to be the most consistent factor that impeded short and long-term recovery in patients with subacute neck pain (Pool, Ostelo, Knol, Bouter & de Vet 2010). Other authors have found that psychological work demands and work hours have an influence on musculoskeletal outcomes and general health (Shannon et al 2001). The perception of poor working conditions has also been shown to be associated with recurrent low back pain, increased medication use and limitation on activities of daily living (Lønnberg, Pedersen & Siersma 2010).

Low back pain of gradual onset showed no significant relation to occupational tasks but has been associated with baseline psychosocial risk factors for back pain (Waddell 2005). Fear-avoidance beliefs related to work have been found to be important determinants in musculoskeletal disability (Denison et al 2004; Lakke et al 2009). Psychological distress and dissatisfaction with the support from work colleagues/management have an adverse effect on the prognosis of back and neck

problems (Macfarlane, Hunt & Silman 2000; Shannon et al 2001; Michaelson et al 2004; Côté et al 2008; Keeley et al 2008). Having lower pay and poor job satisfaction has been shown to be a predictor of outcome for work-related injuries (Koleck et al 2006). Psychosocial factors are associated with the development of chronic neck and back pain and sick leave (Linton 2000). Moreover, patients who changed their job and took more exercise were more likely to improve than workers who remained in occupations with working conditions that were perceived as poor (Carroll et al 2009).

4. **Psychological and Social Factors (potentially modifiable)**

Depression has been cited as an emotional predictor for a less favourable outcome in back pain (Pincus, Burton, Vogel & Field 2002; Carroll, Cassidy & Côté 2004; Dunn & Croft 2004; Hansson et al 2006; Koleck et al 2006; Hill et al 2007; Mallen et al 2007; Henschke et al 2008; Keeley et al 2008; Jensen et al 2009; Lindell et al 2010). Depression has also been found to be a strong and independent predictor for both back and neck pain (Carroll et al 2004; Hill et al 2007). Some authors divide patients into dysfunctional (high pain severity, disability and affective distress), interpersonally distressed (low levels of social support) and adaptive copers (people who successfully adjust to chronic pain by having lower pain levels, better coping strategies, a more positive outlook and less comorbidities) (Bergström et al 2011a; Bergström et al 2011b). Similarly, anxiety and fear-avoidance beliefs relating to work have an adverse effect on prognosis (Denison et al 2004; Hill et al 2007; Mallen et al 2007; Keeley et al 2008). Catastophizing also has a bearing on treatment outcome in some back and neck pain patients (Pincus et al 2002; Hill et al 2007; Schellingerhout et al 2008).

Active coping strategies, with the patient taking measures to control and manage pain themselves have been associated with better outcomes compared to patients who adopt a more passive approach, transferring pain management to an outside source or those who allow others areas of their life to be adversely affected by pain (Pincus et al 2002; Mercado, Carroll, Cassidy & Côté 2005; Mallen et al 2007). However, these authors did find that neck pain patients who had high pain intensity on presentation, had an optimistic attitude and few other symptoms were more likely to improve whereas back pain patients who were likely to improve had higher levels of pain intensity and distress at presentation (Michaelson et al 2004).

5. **Societal Factors (Potentially modifiable but not by the individual)**

Patients with good social support from friends and family have been found to have better control over their pain (Koleck et al 2006, Bergström et al 2011a). Patients who had a perceived risk of chronicity, with poor self-efficacy, poor recovery beliefs, fear avoidance and poor social support are more likely to have a poor prognosis for both back and neck pain (Denison et al 2004; Hill et al 2007; Mallen et al 2007; Foster, Thomas, Bishop, Dunn & Main 2010). These findings are supported by the work of Leboeuf-Yde et al (2004), Enthoven et al (2006), Bergström et al (2010) who also found that these patients were more likely to claim incapacity benefits. Interestingly these authors reported that patients who did not have to pay for their treatment had a less favourable outcome than those who paid for the treatment out of their own pocket or had partial reimbursement (Leboeuf-Yde et al 2009). Back pain patients seeking compensation have been associated with longer recovery times (Hansson et al 2006; Hill et al 2007; Henschke et al 2008, Keeley et al 2008; Jensen et al 2009; Lindell et al 2010) .

Most of the studies reporting predictors of outcome have been done in countries where there is social provision of health care and generous support for those incapacitated by their back or neck pain. Despite this, a recent study of the Nordic subpopulation found the introduction of psychological questionnaires did not prove helpful in predicting patients who would improve under chiropractic care in routine chiropractic practice (Leboeuf-Yde et al 2009). However, these questionnaires did help to identify the dysfunctional and distressed individuals who were more likely to go on sick leave although these findings were limited to male, blue-collar workers (Bergström et al 2011a). Patients attending chiropractic clinics in the United Kingdom, usually in the private sector, display few of the psychosocial predictors associated with poor outcome (Langworthy & Breen 2007).

6. **Genetic Factors (non-modifiable)**

Unfortunately genetic factors have not been well studied in musculoskeletal pain research as prognostic factors (Côté et al 2008; Carroll et al 2009).

7. **Health Behaviours (modifiable)**

Enthoven et al (2006) found higher exercise levels prior to an episode of back pain was associated with a more favourable outcome. Hush, Michaleff, Maher & Refshauge (2009) found patients who took frequent exercise were more likely to respond favourably if they were neck pain patients compared to those with low levels of exercise. However, it has been found that regular cycling has an adverse effect on the prognosis for neck pain, probably because of prolonged periods of neck extension (Carroll et al 2009). Although several authors have found that exercise levels prior to an episode of back or neck pain is associated with a more favourable outcome

(Enthoven et al 2006; Carroll et al 2009; Hush et al 2009), some authors have found that there is no evidence for trunk endurance and low back pain, inconclusive evidence for trunk strength or range of motion and back pain or physical capacity measures for predicting the outcome in an episode of back or neck/shoulder pain (Michaelson et al 2004; Bousema et al 2007; Hamberg-van Reenen et al 2007). A patient with good range of motion, good strength or endurance in their trunk muscles may not experience an improvement in pain severity or in activities of daily living, once again calling into question the usefulness of physical parameters for prediction of treatment outcome (Bolton 1994; Deyo et al 1994; Michaelson et al 2004).

1.5 Limitations of outcome predictor studies

The studies of predictors often look for an association between a limited number of predictors with an outcome although some authors have argued this is a strength in study design as too many variables can result in one or two predictors being incorrectly classified as statistically significant (Hancock et al 2009). Very few predictors from the derivation stage of research have been validated on different populations to see if the predictors of one study are applicable to different populations in new settings. One of the difficulties with these predictor studies is that in assessing treatment predictors, authors have not used the same outcome measures and have confined themselves to a limited outcome such as disability, return to work, pain reduction or self-perceived recovery. Moreover, there is little consensus in the literature as to what constitutes a "successful" outcome (Maughan & Lewis 2010) with different researchers using different measures of "improvement". Finding a suitable outcome measure to determine improvement that is meaningful to both the patient and clinician in chiropractic practice and methods of determining improvement that is meaningful to back and neck pain patients is discussed later in this literature review.

The Michaelson et al study (2004) was conducted on chronic back and neck pain patients who had to have had pain for at least six months. They used what they described as an OS-index, which looked at associated somatic and psychosomatic symptoms (e.g. dizziness, insomnia, depressive symptoms, tiredness), physical parameters such as neck and back endurance, a Visual Analogue Scale (VAS) for pain intensity, the Beck Depression Inventory, the Multidimensional Pain Inventory (MPI), the Optimism Index and the Sociability Index from Edwards Personal Preference Scale. These authors found average pain intensity (VAS) and MPI pain severity were predictive of outcome in back pain whereas a high optimism index, a low sociability scale score, younger patients and fewer associated symptoms were associated with improvement at one year in their neck pain cohort.

Ozegovic et al (2009) used 'return to work expectation' and 'self-perceived recovery' finding patients who expected to return to work reported a global recovery forty two percent more quickly than those who did not have a positive expectation. Bot et al (2005) used "recovery", asking patients whether their symptoms still 'bothered' them, a 11-point numerical rating scale (NRS) for pain and a disease-specific disability questionnaire which the authors developed as a hybrid from several validated disability questionnaires, making comparison with other studies difficult as the authors provided few details on the domains of their disability questionnaire. Tseng et al (2006) used immediate response to cervical manipulation as an outcome and found that patients having a low score (<11.5) on the Neck Disability Index (NDI), bilateral neck pain, not performing sedentary work (>5 hours a day), feeling better when moving their neck, not feeling worse on neck extension and not having radicular pain with their cervical spondylosis where more likely to derive immediate relief from neck manipulation. Fear of neck movement has been shown to consistently impede short and long-term recovery whereas repetitive arm movements have been associated with

patients reporting pain radiating into the upper limbs (Macfarlane et al 2000; Pool et al 2010).

 Skargren & Öberg (2008) looked at disability at one year following chiropractic or physiotherapy treatment and found pain duration (more than 1 month at presentation), the initial score on the Oswestry Disability Index (23 or more) (ODI), less positive treatment expectation (four point scale), the number of painful areas (more than 1 area-neck, thoracic, low back) and the patients' perception of well-being (six point scale of well-being and a VAS) affected outcome. Whether the patient received chiropractic treatment or physiotherapy did not affect the outcome for the twelve percent of this cohort who had the unfavourable predictors at the outset of the trial. May, Gardiner, Young & Klaber-Moffett (2008) studied the functional outcome (at least a fifty percent reduction in score on Roland Morris Disability Questionnaire [RMDQ]) in patients with acute and chronic neck and back pain receiving McKenzie treatment and found duration of pain at the time of presentation was the strongest predictor of success with pain centralising to the spine from peripheral areas and localised low back rather than accompanying neck pain as weaker predictors.

Moreover, Wand, McAuley, Marston & De Souza (2009) questioned whether one should use baseline variables to predict outcome at all? Their study on low back pain patients found that baseline variables could predict long-term pain only. However, the six-week profile was useful for predicting both long-term pain and disability. These authors used a wide array of outcome measures, the RMDQ, the Spielberger State-trait Anxiety Inventory, the Zung Depression Questionnaire, the Modified Somatic Perception Questionnaire and the Short Form-36 (SF-36). The subjects of this study were non-specific low back pain patients attending the physiotherapy or Accident and Emergency Department of a London Hospital. Some of the outcome measures used in this study were surprising as one associates the

choice of some of these outcome measures with chronic patients with complex psychosocial complications rather than patients presenting with non-specific low back pain. This might explain why these authors, unlike other studies, did not find baseline variables useful as treatment predictors. Interestingly, nearly half the cohort did not complete follow-up questionnaires in this study which might have affected the outcomes.

Ignoring baseline data and using data at six weeks after initial presentation would present problems in routine clinical practice as most patients would have been discharged six weeks after their initial consultation. Some authors have found that outcome at the fourth visit was predictive of outcome at three months but a subsequent study by the same author failed to reproduce these findings (Leboeuf-Yde et al 2004; Leboeuf-Yde et al 2009). The wide diversity of choice of outcome measures could account for why researchers have found different predictors of treatment outcome as these questionnaires often measure different domains and vary in their ability to detect change in different study populations.

1.6 Outcome measures

Whatever the condition, there is agreement, as illustrated by the work of the European Agency for the Evaluation of Medicinal Products and the US Food and Drugs Administration, that the patient's voice should be heard using the most psychometrically-sound measures available since the patient's perspective on treatment outcome is crucial (Revicki 2007). Bodies such as PROMIS (Patient-Reported Outcomes Measurement System) have arisen to ensure researchers develop reliable and valid patient-reported outcomes and although the use of such outcomes in clinical trials is more common, this is by no means always the case, meaning valuable information regarding patients' attitudes, beliefs and coping strategies is being missed (Scoggins & Patrick 2009). Variability in outcome measures across trials makes

it difficult to compare treatment effects as they do not all measure the same domains.

Consequently the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) recommended that six core outcomes should be included in pain trials: pain, physical functioning, emotional functioning, patient improvement ratings/satisfaction with care, symptoms/adverse effects and patient disposition (Turk et al 2003). These outcome measures, originally designed for clinical trials, are now experiencing wider use in clinical practice to assess treatment outcomes and provide information to both patients and to the treating clinician. The variables tested in predictor studies are usually derived from the baseline data of outcome measures together with demographic and clinical characteristics of the patient.

Chiropractors in clinical practice report that the results of clinical trials are often at odds with the reality of clinical practice (Bolton 2003). There are many reasons why this happens but one of them could be the choice of outcome measure (Lemieux, Beaton, Hogg-Johnson, Bordeleau & Goodwin 2007; Revicki, Hays, Cella & Sloan 2008). One of the most widely-used patient-report outcome measures used in clinical trials is the SF-36 (Scoggins & Patrick 2009; Dawson, Doll, Fitzpatrick, Jenkinson & Carr 2010). Although there are many condition-specific outcome measures, generic measures such as the SF-36 remain popular (Khorsan, Coulter, Hawk & Choate 2008; Scoggins & Patrick 2009). The advantages are that one can compare patients in different disease-groups, one might detect unexpected effects, one can often look at cost-effectiveness, that generic measures tend to be shorter than disease-specific instruments and outcome measures such as the SF-36 remove the need to find or develop an outcome measure for every condition under investigation (Khorsan et al 2008). However, years of validation studies have shown poor correlation between reported function and measured performance using the SF-36, calling into question its use in

detecting clinically meaningful improvements in back and neck pain patients (Samsa et al 1999).

Disease-specific instruments (instruments specific to a condition) reduce patient burden and tend to be more responsive to change in patient health status compared to generic measures (Fletcher et al 1992). The RMDQ (Roland & Morris 1983) and the ODI (Fairbank, Couper, Davies & O'Brien 1980) remain the most widely used outcome measures in low back pain research and have been translated into many languages (Peat 2004). This author states that the RMDQ and the ODI are recommended for evaluating patient self-reported low back pain-specific functional limitation. However, the RMDQ does contain some items that do not relate to function and the ODI has many versions that are in use. Some patients object to answering questions about their sex life which results in missing data for the ODI but patients also fail to complete fifteen percent of questions on the RMDQ. The RMDQ essentially evaluates function and looks as some aspects of illness behaviour whereas the ODI only assesses pain intensity and disability. The most widely used outcome measure for neck pain is the NDI (Vernon & Mior 1991), modified from the ODI but this outcome measure, like the ODI, only evaluates pain intensity and disability, missing several important domains of the BPS model for neck pain.

The measurement of pain itself is complex and there are various methods of assessing the distribution, duration and nature. However, the most widely measured domain is pain intensity. The VAS is a ten-centimetre horizontal line anchored with labels of "no pain" and "worst possible pain" where patients mark pain intensity with a vertical line on the scale, the Numerical Rating Scale (NRS) asks patients to rate their pain intensity on an 11-point scale (0=no pain and 10=worst possible pain) and the Verbal Rating Scale (VRS) where adjectives or descriptors are used in place of numbers (Coons 2008). For back pain patients, the NRS

has been shown to be most responsive to change, especially if patients are asked to rate their average pain rather than the worst pain they have experienced (Bolton & Wilkinson 1998). Pain drawings have been used for nearly seventy years and give valuable information about pain distribution (Ohnmeiss 2000; Lacey, Lewis, Jordan, Jinks & Sim 2005). They have been used for psychological evaluation such as hypochondriacal tendencies but cannot be relied upon as a measure of psychopathology (Udén, Aström & Bergenudd 1988, Jensen & Karoly 1992; Voorhies, Jiang & Thomas 2007).

It has been recognised that pain drawings, quantitative and qualitative pain scales and disability scales yield findings that are often at odds with objective clinical findings which suggests that there is a psychological aspect to the pain experience, especially in chronic pain patients (Jensen & Karoly 1992; Voorhies et al 2007). A number of questionnaires have been developed that examine the emotional factors (distress), personality traits, cognitive factors (coping strategies) and behavioural factors (illness behaviour) (Bolton 1993; Bolton & Breen 1999). To assess all the domains in the BPS model of the back/neck pain experience, one would have to use questionnaires for qualitative pain, quantitative pain, disability, distress, anxiety, depression, coping strategies and illness behaviour for each patient. This would be impractical in terms of time and complexity as these questionnaires are often lengthy, difficult to interpret, can be intrusive in the nature of the questions and many of these outcome measures can only be used under licence. Only using one or two measures would mean one could not evaluate the complexity of the compliant, arriving at conclusions that could be neither relevant nor meaningful. This led to the development of the Bournemouth Questionnaire (BQ), examining all the commonly-measured domains of the BPS model for back and neck pain patients (Bolton & Breen 1999; Bolton & Humphreys 2002).

1.7 Bournemouth Questionnaire

The aim of the study developing the BQ was to develop a clinically useful, multi-aspect outcome measure that was based on the BPS model of back pain, that would be brief, with each domain being measured on a single-item scale, that would be useful for the ambulatory patients that seek the services of chiropractors, that would be easy and quick to use in a clinical or research setting, acceptable to patients, practitioners and researchers as well as being reliable, valid and responsive to clinically significant change for both back and neck patients (Bolton & Breen 1999; Bolton & Humphreys 2002; Dawson et al 2010). This outcome measure had to reflect back and neck pain as illnesses rather than diseases and focus on management rather than treatment (Bolton & Breen 1999; Bolton & Humphreys 2002). The items included in the final questionnaire, designed for back pain patients, were pain intensity, daily functional activity and social activity, the affective dimensions of anxiety and depression as well as the cognitive/behavioural dimensions of fear-avoidance beliefs and self-efficacy beliefs of pain control (Bolton & Breen 1999). Validity and reliability of the BQ was tested against the Chronic Pain Questionnaire, the RMDQ, the Modified Somatic Pain Questionnaire, the Fear Avoidance Beliefs Questionnaire (FABQ) and the Pain Locus of Control Questionnaire and all items were statistically significant (Bolton & Breen 1999). It was demonstrated that a change in total scores greater than four-and-a-half points was indicative of real change (Bolton & Breen 1999).

Mindful of the similarities between back and neck pain, the BQ was modified and found to be valid, reliable and responsive for neck pain patients (Bolton & Humphreys 2002). Apart from locus of pain control, the validity of the BQ for neck pain patients was good compared to the NDI and the Copenhagen Neck Disability Index (Bolton & Humphreys 2002; Gay,

Madson & Cieslak 2007). Similarly, the BQ was found to be considerably more responsive to change than the NDI and the Copenhagen Neck Disability Function Scale for neck pain patients (Bolton & Humphreys 2002). The BQ demonstrated reasonably good reliability and a total change score greater than twelve points was indicative of real change in neck pain patients (Bolton & Humphreys 2002).

The BQ has been translated and validated in French, Dutch and Danish in the peer-reviewed literature, into German at the University of Zurich by one of the original authors and into Swedish and Spanish for post-graduate MSc dissertations (Malmqvist 2001; King 2002; Hartvigsen et al 2005; Khorsan et al 2008; Martel, Dugas, Lalond & Descarreaux 2009; Schmitt et al 2009). However, like other outcome measures, no cross-cultural validation of the BQ exists (Schellingerhout et al 2011). It is being used at undergraduate teaching clinics and more widely by clinicians in practice.

In addition to the seven core questions on the BQ, the Pre-treatment BQ (Appendix II) is preceded by a number of clinical and demographic questions about the nature and course of the pain experience. The Post-treatment BQ (Appendix IV) contains a Patient Global Improvement Scale (PGIS) which was extrapolated from the Clinical Global Impression of Change (CGIC), the most common version of which includes a seven-point Likert scale used to assess global improvement in psychiatry but widely-used in musculoskeletal research. Scales usually range from 'very much better' to 'very much worse' (Khorsan et al 2008). The PGIS has been used extensively as an external criterion in outcome studies to 'anchor' change scores in outcome measures of improvement or deterioration (responsiveness) (Hägg, Fritzell & Nordwall 2003; Hurst & Bolton 2004; Mannion et al 2009; Krebs et al 2010, Schäfer, Hall, Müller & Briffa 2011). The post-treatment BQ also has questions about sick leave, work status, satisfaction with treatment received and whether or not treatment

expectations have been met (Bolton & Breen 1999; Bolton & Humphreys 2002). Measurement of patient satisfaction has become more prevalent and instruments to do this have been validated (Beattie, Nelson & Murphy 2011).

1.8 Use of Outcome Measures in Clinical Practice

Measures originally designed for assessing treatment effectiveness in clinical trials are now being used routinely by the NHS since April 2009 to assess patient perspective and treatment outcomes for hip and knee replacement, inguinal hernia repair and varicose vein surgery (Dawson et al 2010). Systematic use of Patient Report Outcome Measures (PROMS) may result in improved patient outcomes providing patient-centred information facilitating improved communication between doctor and patient. Patients also feel that their view/perspective of their health and well-being are being taken into consideration (Dawson et al 2010). In a recent study back, neck and shoulder patients were able to access spinal manipulation (osteopathy) and acupuncture on the NHS in a general practitioner's surgery. The BQ was one of three outcome measures used to assess patient progress and all seven domains of the BQ were found to produce highly statistically significant change or improvement for this cohort of patients, assessing the BPS aspects of the pain experience whereas the other outcome measures (EuroQoL-5D, medication use, physical activity and general well-being) failed to demonstrate statistically significant change (Cheshire, Polley, Peters & Ridge 2011).

Apart from condition-specific and generic measures, in recent years a third category of patient outcomes has arisen, the patient-specific instrument which is neither generic nor condition-specific. In this type of outcome measure the patient generates their own often unique items which can offer valuable information about changes in quality of life about an

individual patient and what is important to them (Jolles, Buchbinder & Beaton 2005). An example of such an outcome measure is the Measure Yourself Medical Outcome Profile (MYMOP), in limited use by some chiropractors linked with the College of Chiropractors (Paterson 1996). This outcome measure has been shown to be more sensitive to change than the SF-36 and has the added bonus of improving patient-practitioner communication but there is little evidence that this type of outcome measure shows changes over time (Paterson 1996; Paterson 2004). Moreover, as each patient's items are self-determined by the individual patient and are often unique, it is not possible to use this kind of outcome measure to compare groups of patients over time and would not be an appropriate outcome measure for a study such as the Cheshire et al (2011) paper.

1.9 Statistical versus Clinical Significance

Traditionally, evaluation of treatment interventions using subjective outcome measures has been based on the statistical significance of the difference between group mean values, indicating that observed group changes are large enough not to have been caused purely by chance (Kamath & Dueck 2005). However, information is required from sensitive outcome measures on the proportion of patients undergoing a clinically important improvement for the results of a study to be meaningful to a practising clinician trying to determine if the reported intervention would be of use to his/her patient (Jacobson, Follette & Revenstorf 1984; Jacobson & Traux 1991; Turk 2000; Bolton 2004). Statistical analysis usually reports group differences between different treatment interventions, treatment with placebo, experimental treatment with standard care, experimental treatment with a waiting list, experimental care with those who refused treatment or changes in a single group before and after treatment (Turk 2000).

As the sample size increases, the statistical result necessary to achieve statistical significance decreases. In other words, if one finds a small difference between two groups, it is likely that one will achieve a statistically significant result simply by introducing subjects to the study (Jacobson et al 1984; Turk 2000; Crosby, Kolotkin & Williams 2003). Furthermore, statistical comparisons, based on group means produce no data on the variability of the data, make assumptions about the distribution of data that may or may not be true and offer little or no information on the individual patient treatment effect (Jacobson et al 1984; Ottenbacher, Johnson & Hojem 1988; Wyrwich & Wolinsky 2000). Statistical significance does not indicate the proportion of individuals in the group who have achieved a clinically meaningful change from the treatment intervention (Maughan & Lewis 2010). Randomised controlled trials, albeit the gold standard of treatment efficacy, only give data on group averages. The predictive value of the prognostic factors is limited, probably because so many prognostic factors interact with treatment outcomes (Hagen et al 2005).

The threshold for the determination of clinical improvement has been described variously in the literature: Minimal Important Change (MIC), Minimal Important Difference (MID), Minimal Clinically Important Difference (MCID) and Minimal Detectable Change (MDC) (Houweling 2010). This entity was first defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management" (Jaeschke, Singer & Guyatt 1989). Subsequent changes in this definition have appeared in the literature in subsequent years: "the smallest difference in a score that is considered to be worthwhile or important" (Hays & Woolley 2000) or "the mean score for patients with an optimal result minus the mean score for a group with suboptimal results" (Quinn & Wells 1998). Recently an international panel of experts stated that a thirty percent

change from baseline may be considered clinically meaningful when comparing before and after measures for individual patients (Ostello et al 2008). However, some scepticism exists in the literature as to whether we will ever be able to reach one MCID for an instrument (Beaton, Boers & Wells 2002). The study design and the population sample are bound to have an influence in determining the proportion of patients who will have improved (Houweling 2010).

1.10 Methods of Determining Clinically Significant Improvement

Broadly speaking two approaches have emerged in the literature for determining the MCID: the anchor-based and distribution-based methods (Copay, Subach, Glassman, Polly & Schuler 2007; Houweling 2010). Distribution-based methods measure change alone whereas anchor-based methods measure clinically meaningful change (Maughan & Lewis 2010). Anchor-based methods are based on comparing a change in score of an outcome measure with an external criterion or "anchor" (Wyrwich &Wolinsky 2000; Copay et al 2007; Houweling 2010). The most commonly used anchor is a PGIS where patients rate themselves as better, unchanged or worse. This kind of anchor is not ideal as it is not objective and as such is not a "gold-standard" (Copay et al 2008). There are limitations in using a PGIS as there could be recall bias where patients underestimate their initial state and provide retrospective estimates of change based on their current health status although PGISs have been shown to be very sensitive to change, both negative and positive (Wyrwich & Wolinsky 2000; Copay et al 2007).

Anchor-based methods

There are four anchor-based methods. The first method is "within-patient" score change, where the MCID is defined as the mean change score in the "much-improved" group on the

PGIS but the selection of this group of patients is arbitrary (Copay et al 2007). The "between-patient" score change is where the MCID is the difference between the change score between the "better" and "unchanged" patients. Once again, the choice of two adjacent levels of the PGIS is arbitrary (Copay et al 2007). The third method is where patients compare themselves to other patients. This method has not been widely used and is fraught with difficulty as patients generally consider themselves better off than the person to whom they are comparing themselves (Guyatt & Jaeschke 1997) and they judge a small difference to be more meaningful when they felt better rather than worse compared to the person to whom they were comparing themselves (Wells et al 1993). Furthermore, disparity between what constitutes clinically meaningful change using this anchor-based method compared to other anchor-based methods has been found, questioning the accuracy of this method (Wyrwich & Wolinsky 2000). It is likely that the noise and potential measurement error introduced by each patient's perception of both their own health status and that by exhibited by their paired-partner may obscure the underlying between-person differences (Wyrwich & Wolinsky 2000).

The fourth and most widely used method is the determination of the MCID using a Receiver Operating Characteristic (ROC) curve to find the most accurate cut-off point that maximises the score between sensitivity (true-positives) with specificity (true negatives) (Copay et al 2007, Houweling 2010; Maughan & Lewis 2010). Sensitivity is the proportion of patients who report improvement on the external criterion (usually PGIS) and whose change in score on the outcome measure is above the threshold for the MCID value. Specificity is the proportion of patients who do not report an improvement on the external criterion and whose outcome measure score change is below the MCID threshold value (Copay et al 2007). It is usual for the ROC curve-derived MCID to be the change score on the outcome

measure that separates the "somewhat better" patients from the "about the same" patients on the external criterion with equal specificity and sensitivity (Copay et al 2008). The point on the ROC curve that is closest to the upper left corner of the figure is taken as the MCID, the change score with the minimum amount of misclassification (Maughan & Lewis 2010; Terwee et al 2010). The area under the ROC curve can be interpreted as the probability of identifying improvers from non-improvers and the greater the area under the curve indicates the instrument's accuracy (Maughan & Lewis 2010).

MCID studies search for a unique threshold value but different methods of calculating clinical significance produce a variety of MCID values. The anchor-based methods will produce different MCID depending on the external criterion scale and arbitrary selection or grouping of scale levels (Copay et al 2007). It is possible that looking for the MCID between "unchanged" and "slightly better" might be so small as to lack statistical significance and fall within the boundaries of measurement error (Copay et al 2007). ROC curves require dichotomous variables (improved/not-improved) and this necessitates an arbitrary selection or grouping of subjects (Copay et al 2007). Recent research has found that ROC analysis should include the entire cohort rather than including only the patients around the dichotomisation point as it yields higher sensitivity, specificity and narrower confidence intervals (Turner et al 2009; Turner et al 2010). The ROC method, complemented by a graph of the anchor-based MCID distribution provides all the necessary information when MCID values are applied to individual patients (De Vet et al 2010). However, other difficulties with the anchor-based method are patient recall bias, low or unknown reliability/validity of the external criterion and low correlation between the external criterion and the actual change score on the outcome measure (Crosby et al 2003; Kamath & Dueck 2005; Lauridsen, Hartvigsen, Manniche, Korsholon & Grunnet-Nilsson 2006).

Estimates of the MCID using anchor-based methods depend on the external criterion, the direction of change and baseline values. Previous research has shown that patients with high baseline scores require higher thresholds to demonstrate clinical significance and acute patients seem to require a higher threshold than chronic patients (Houweling 2010). As anchor-based methods do not take the standard deviation of the group into consideration, this could lead to an underestimation of the number of subjects classified as improved (Houweling 2010). For these reasons, it is unwise to accept the results of this methodology without comparing the results of at least one of the distribution-based methods (Kamath & Dueck 2005). However, as distribution-based methods lack information as to whether the observed changes are minimally important, anchor-based methods are the preferred methodology with distribution-based methods providing the necessary supportive evidence (De Vet et al 2010).

Jordan, Dunn, Lewis & Croft (2006) argued that the anchor-based method, using the RMDQ for back pain patients had very good specificity but lacked sensitivity and for that reason argued distribution methods should be preferred as they took into account the reliability of the outcome measure and produced a cut-off that was beyond measurement error although they did concede patients with less severe disease cannot improve using this method. All other authors who have studied clinical significance calculations agree that anchor-based methodology gives a better indication of clinical significance and current anchor-based research seems to use the ROC curve analysis to determine cut-off values with the best sensitivity and specificity (Copay et al 2007; Revicki et al 2008; Terwee et al 2010).

Distribution-based Methods

Distribution-based methodology depends on the distribution of the population sample under scrutiny (Houweling 2010). An advantage of the distribution-based methods is that they provide a way of establishing change beyond random variation (Kamath & Dueck 2005; Jordan et al 2006). Distribution-based methods can only define a minimum value below which a change score on an outcome measure may likely be because of measurement error and as such, these methods only provide a minimum detectable change, which indicates nothing about clinical importance, or in other words, minimal detectable change rather than minimal important change (Gatchel & Mayer 2010; Terwee et al 2010).

In 1987 effect size standards for classifying individual patient change were introduced (Testa 1987). This author modified the Kazis formula for group change over time by calculating the individual effect size as the individual score at follow-up less the individual baseline score divided by the standard deviation of the group at baseline (Testa 1987; Wyrwich & Wolinsky 2000). Modifying Cohen's benchmarks for group effect size of 0.2 for small group change, 0.5 for moderate group change and 0.8 for a large group change, Testa proposed 0.2, 0.6 and 1.0 respectively for individual effect sizes (Testa 1987; Wyrwich & Wolinsky 2000). As Cohen's benchmarks were not designed for subjective outcome measures and standard deviations are sample specific, effect size estimates can vary widely among samples taken from the same population (Samsa et al 1999; Wyrwich & Wolinsky 2000). In more recent years benchmarks of 0.2, 0.33 and 0.5 have been suggested for dichotomising improved from not improved patients (Sloan, Cella & Hays 2005; De Vet et al 2006, Copay et al 2007; Pickard, Neary & Cella 2007 ; Puhan, Frey, Buchi & Schunemann 2008; Revicki et al 2008).

Half a standard deviation (0.5SD) of the baseline score was suggested as this is equal to an effect size of 0.5, generally considered to be a moderate effect (Norman, Sloan & Wyrwich 2003; Eton et al 2004; Kamath & Dueck 2005; Lemieux et al 2007; Revicki et al 2008; Turner et al 2010). Although this magnitude of change is certainly clinically significant and meaningful, it is not necessarily minimal (Revicki et al 2008). It is necessary to acquire information as to whether the observed change is important to patients or clinicians. MCIDs as small as 0.25 have been demonstrated using distribution-based methods, which are simply a way of expressing observed change in a standardised metric, but could be clinically meaningless (Webb et al 2003). It is possible that in certain settings an effect size of 0.5 or 0.5SD might be setting criteria for success beyond that which is achievable in a given treatment as the threshold for clinical significance could be lower (Revicki et al 2008).

The Standard Error of Measurement (SEM) relates the reliability coefficient of the outcome measure to the group baseline standard deviation (baseline group score standard deviation multiplied by the square root of one minus the reliability coefficient of the outcome measure) (Houweling 2010). This method takes account of the precision of the instrument (outcome measure) which is an aspect not taken into account in anchor-based methodologies (Rejas, Pardo & Ruiz 2008). Although 1, 2 and 2.77 SEM have been used as cut-off points for dichotomising patients, many authors have consistently found one SEM corresponds to previously established MCID standards (Wyrwich & Wolinsky 2000; Copay et al 2007; Copay et al 2008; Rejas et al 2008; Revicki et al 2008; Terwee et al 2010; Turner et al 2010). The SEM is fairly sample independent and consistently gives similar results to 0.5SD (Houweling 2010). One SEM has been found to relate well to results using the PGIS in anchor-based methodology (Jordan et al 2006). With a reliability factor of 0.75 for an outcome measure, one SEM is equivalent to 0.5SD (Terwee et al 2010).

The Reliable Change Index (RCI) first proposed by Jacobson et al (1984) was the first attempt to calculate clinical significance. Modified by Christensen and Mendoza in 1986, this index is a derivative of the effect size and the SEM (The mean post-score minus the mean pre-score divided by the square root of two multiplied by the squared standard error of measurement) (Christensen & Mendoza 1986; Houweling 2010). Also taking precision of the outcome measure into account, this formula yields larger values for the denominator and thus RCI values are smaller than SEM-based values. A cut-off of 1.96 was used to decide if the patient was classified as improved/not improved (Crosby et al 2003). This formula was further modified for individuals rather than groups of subjects with a less rigorous cut-off of 1.65 but although recommended for its statistical properties, the RCI has poor sensitivity to classify improved patients, being too conservative and as a result, it has appeared very infrequently in current literature, although it corresponds well to clinician assessment, generally regarded as the least credible of the anchor-based methods (Hageman & Arrindell 1999; Norman et al 2003; Jordan et al 2006).

Evidence from various studies suggest half standard deviation is a meaningful to patients (Revicki et al 2008, Houweling 2010). It has been often described as a good proxy if an anchor-based methodology is not possible although this has been criticized as it does classify patients as false positives meaning changes below the mean are ignored (Lemieux et al 2007). The SEM is hampered by yielding a high threshold if the reliability coefficient of the outcome measure is not good and it is questionable as to whether it is appropriate to use the reliability coefficient for the determination of patients who have improved as this coefficient has usually been calculated for a completely different sample of patients than the one under scrutiny (Eton et al 2004; Houweling 2010). Proponents of particular distribution-based methods vigorously defend their chosen method such as the SEM, half standard

deviation or effect size but there is no compelling evidence for supporting any one method as a standard and it was not until Copay et al (2008) that any author attempted to systematically compare all the methods on the same patient sample. Several authors have found that the 0.5 SD and 1 SEM corresponded closest to the anchor-based values for small change and the RCI compared best with clinician rating of improvement, leading them to suggest that distribution-based methods should only be used as temporary substitutes for anchor-based methods (Rejas et al 2008; Revicki et al 2008; Turner et al 2010). Distribution-based methods emphasize statistical significance and do not necessarily address clinical significance or the MCID with the MCID being sample specific, depending on the variability of scores in the studied sample (Copay et al 2007).

1.12 AIMS AND OBJECTIVES

This literature review has shown considerable research into predictors within the field of musculoskeletal pain and disability. In spite of this research, questions still remain as to whether or not there are robust predictors in back and neck pain patients and, if so, what these predictors are. Given the increasing interest in predicting the outcome in patients, this study was set up to investigate patients undergoing chiropractic treatment in a practice in Bristol. This was a pragmatic study based on outcome measured by patient self-reported 'improvement' and evaluating a number of potential predictors in back and neck pain patients.

The objectives of this study were:

1. To describe the clinical and demographic characteristics (potential predictor variables) of patients with back and neck pain (with or without headache) presenting for chiropractic treatment in routine clinical practice.

2. To determine the most reliable and valid method of identifying which patients had improved at the 4th/5th and 10th treatment session (determination of outcome).

3. To determine the variables independently associated with improvement in back and neck pain (with or without headache) patients at the 4th/5th and 10th treatment visits.

Chapter 2: Methods

2.1 Patients

Consecutive new patients presenting to a private chiropractic practice completed a 'New Patient' questionnaire (see appendix i) and the self-report pre-treatment BQ (see appendix ii) in reception prior to seeing the chiropractor. All new patients formed part of this study except patients who were not fluent in English, patients too elderly or frail to comprehend the questionnaire, patients under sixteen years of age and the occasional patient declaring that the 'questions were not relevant to their condition'.

2.2 Questionnaires

The questionnaires contained a list of demographic and clinical questions concerning the patient's presenting complaint. There was also a section where patients were asked to shade their area/s of pain on a diagram of the homunculus from the posterior and anterior aspect. Two methods were used to evaluate the markings patients made on the pain diagrams:

1. The method of penalty points (yes/no) where patients shaded pain in a sporadic, non-physiological pattern, indicated pain outside the body or used additional words or symbols to describe the pain (Ohnmeiss 2000).

2. The grid method (Ohnmeiss 2000) superimposing a transparency with a number of squares covering the image of the homunculus (see appendix iii). A separate score was recorded for the anterior and posterior aspect of the body. Each square on the grid that was shaded counted as one point with a maximum score of 52 points.

In addition to the 'New Patient' Questionnaire, the pre-treatment BQ was used. This questionnaire has been validated for chiropractic patients in a research and clinical setting

and has been shown to be valid, reliable, responsive to change and has been validated in several other languages (Bolton & Breen 1999; Bolton & Humphreys 2002; Hartvigsen et al 2005; Martell et al 2009; Schmitt et al 2009). This questionnaire consisted of demographic and clinical questions as well as the seven core BQ questions consisting of the most commonly measured domains of the BPS model of back/neck pain (Bolton & Breen 1999; Bolton & Humphreys 2002), namely pain intensity, disability in activities of daily living (ADL), interference in social life, anxiety, depression, work fear-avoidance and locus of control. Each domain was measured on a numerical rating scale from 0-10, with a total maximum score of 70/70, with a high score on each subscale representing a worse condition (Figure 1).

For the purposes of reassessment on the 4th or 5th treatment and the 10th treatment if the patient was still consulting¸ the post-treatment BQ, which asked the same seven core questions on numerical rating scales giving a total maximum score of 70/70, was used (see appendix iv). In addition there were questions concerning current work status, satisfaction with the care received and whether the results of treatment had matched expectations. In addition to these questions, the post-treatment BQ also contained a PGIS where patients were asked to choose one of seven options from 'no improvement' through to 'a great deal better and a considerable improvement which has made a worthwhile difference' (see appendix iv).

All new patients except those under sixteen years of age, those not fluent in English, patients too frail to complete the questionnaires and patients who felt the questions were not relevant to their condition completed New Patient Questionnaire including a Pain Diagram and the Pre-Treatment Bournemouth Questionnaire

Patients either discharged or self-discharged/lost to follow-up

After patients had undergone 4 or 5 treatments, they completed the Post-Treatment Bournemouth Questionnaire including a Patient Global Improvement Scale in reception prior to leaving the clinic

Patients either discharged or self-discharged/lost to follow-up

During the last 17 months of data collection all patients requiring a more protracted course of spinal manipulation completed the Post-Treatment Bournemouth Questionnaire after treatment 10

Figure 1. Flow chart of patient data collection during the study

2.3 Procedure

Patients completed the 'New Patient' Questionnaire as well as the pre-treatment BQ in the time prior to their initial consultation with the chiropractor. These questionnaires were given to the patients by the receptionist and returned to the reception desk once completed. Following the 4th or 5th chiropractic treatment, when the patient had been formally re-examined as part of their ongoing re-assessment, patients completed the post-treatment BQ in reception prior to their departure. Chiropractic treatment sessions usually consisted of asking the patient how they been since their last visit, an examination of their area of

complaint followed by treatment which could consist of one intervention or a combination of spinal manipulation, mobilisation, stretches, exercises, soft tissue work or dry needling (acupuncture). Many patients in pain develop fear avoidance and become fearful of certain activities. Patients were given advice on remaining active, trying to remain at work even though they may have needed to change their duties, not to sit for too long at any one period and to take plenty of purposeful short walks. This advice and reassurance is in line with current guidelines, the most recent of which was published by NICE (Savingy et al 2009). Since September 2007, patients needing to undergo a more protracted course of treatment were also asked to complete a second post-treatment BQ following re-examination on their 10th treatment visit if they were still consulting.

2.4 Data Management

Data were entered onto an Excel Spreadsheet with each patient being coded with a number. The patient's name was also entered to make it easier to locate the patient's file for the purposes of data clean-up. However, patient anonymity was assured at the time of data analysis by removing the patient's name. The patient's age was recorded as the age at the date of presentation. All positive answers were coded with a value of '1' and negative answers with a value of '2'. Each pain diagram was evaluated by an experienced chiropractor for sporadic, non-physiological markings, markings outside the body or additional symbols/words and given a value of '1' if considered as 'inappropriate marking' and '2' if the markings were unremarkable. Using the grid system, the number of squares shaded on the pain diagram was recorded for the posterior and anterior aspect of the homunculus (maximum value 52). Values for each of the seven core domains of the BQ completed by the patient on a scale from 0-10 were recorded. For questions with more than two option answers, for example, "How do you expect your condition to change in the next two weeks?-

'recover/improve', 'stay the same' or 'get worse'", values of 1, 2 and 3 were given for each answer respectively.

All data were transferred to Statistical Package for Social Scientists (SPSS) version 16. The data were cleaned up by running descriptive statistics and replacing outliers/nonsensical data with the missing value by referring back to the patients file to obtain the correct value for outlier/nonsensical data. For example, on a few occasions the duration of the patient's treatment in days was spread over years. Referring back to the patient's file enabled the researcher to establish the correct dates and determine the correct duration of treatment in days. Dichotomisation of potential predictor variables took place at this stage to facilitate data analysis (Table 1). There was an inherent risk that dichotomisation could result in loss of information and power in data analysis. However, whilst recognising the loss of power, one had to consider interpretability of the results and how to apply this in practice. For a clinician reading results one could say if the patient has "x" or "y" they are more likely to recover whereas with five or six categories in an ordinal scale interpretation of results could pose considerable difficulties and as such, variables were dichotomised, supported by the available literature. Moreover, researchers comparing dichotomous data with continuous measures have found that dichotomous data can approach or even exceed that of a well-defined continuous measure in musculoskeletal research (Anderson 2007).

Patients were dichotomised at the 75th percentile for age. As previously discussed prevalence of back and neck pain increases with age (Andersson 1999; Webb et al 2003, Parsons et al 2003, Parsons et al 2007; Savingy et al 2009; Keijsers et al 2010). Moreover, other authors have stated older patients have a less favourable outcome (Leboeuf-Yde et al 2004; Michaelson et al 2004; Hill et al 2007; Mallen et al 2007, Henschke et al 2008;

Schellingerhout et al 2008). It is known that patients seeking compensation have a less favourable prognosis so patients whose pain was as a result of a specific trauma were dichotomised from those who were not or did not know if their injury was as a result of trauma (Buitenhuis, de Jong, Jaspers & Groothoff 2008; Rasmussen, Leboeuf-Yde, Hestbaek & Manniche 2008, Derebery, Giang, Gatchel, Erickson & Fogarty 2009). The literature states that patients who have a favourable treatment expectation are associated with a better outcome and as a result patients expecting to recover/improve were dichotomised from those who expected to stay the same or get worse (Hall et al 2007; Myers et al 2007). Whilst no clear link between employment status and prognosis has been established, some authors have associated duration of sick leave with prognosis (Michaelson et al 2004). All the baseline variables that were dichotomised can be found in Table 1.

For the purpose of analysis, total BQ scores were used rather than individual core domains giving a score out of a total of 70. Previous authors have analysed the seven domains of the BQ individually (Bolton & Breen 1999; Bolton and Humphreys 2002) and found the BQ taps into different aspects of the same attribute (Bolton & Breen 1999; Bolton & Humphreys 2002). Earlier research has explored methods of categorising improved and not improved patients and although these methods have been refined in recent times, previous work has used the total BQ score (Hurst 2001; Hurst & Bolton 2004). The dichotomisation of the total BQ score was chosen after the researcher had transferred patient data onto a spreadsheet for 500 patients when a clinical feel developed that scores in excess of 30/70 were associated with a less successful treatment outcome. The same clinical impression developed for the pain diagram with patients who shaded 8 or more squares being dichotomised.

Table 1. Categorisation (dichotomisation) of potential predictor variables

Age	≤50 years(75th percentile)/>50.1 years
Gender (male)	Yes/no
Radiating pain to leg or arm	Yes/no
Widespread pain	Yes/no
Trauma	Yes/no or don't know
Similar complaint in the past	Yes/no
Duration of episode	<7 weeks/>7 weeks
Medication use a lot of the time	Yes/no
Treatment expectation	Recover-improve/stay the same or get worse
Work status	Employed/seeking work or retired or working in the home
Job satisfaction	Yes/no
Expecting to work normally in 6 months	Yes/no
Current or previous smoker	Yes/no
Regular alcohol consumption	Yes/no
More or same level of physical activity as peers	Yes/no
Excellent or good general well-being	Yes/no
Number of areas on pain diagram (<8)	Yes/no
Inappropriate signs on pain diagram	Yes/no
Total pre-treatment BQ score max (70/70)	30/70 or greater

2.5.1 Data Analysis

BACK AND/OR LEG PAIN PATIENTS

Data from these patients were analysed in six stages:

1. Descriptive statistics of the study population

2. Different methods of categorising improved/not improved patients

3. Associations between baseline variables that could be potential predictors

4. Univariate analysis of the baseline variables and interactions between the variables

5. Multivariable regression analysis

6. Properties of the prediction model

STAGE 1

Descriptive Statistics

Descriptive statistics were used to analyse the data: Categorical data were described in numbers and percentages, and continuous data were described as means ± standard deviation.

STAGE 2

1. **Assessment of improved/not improved patients**

Three methods were used to determine patient outcome in terms of improved/not improved.

   a. **Direct Method**

   Using the PGIS in the post-treatment BQ, patients marking option 6 or 7, namely 'a definite improvement' or a 'considerable improvement' were classified as improvers and those choosing options 1-5 being classified as non-improvers. This robust cut-off has demonstrated good correlation with other methods of classifying improved patients (Hurst 2001; Hurst & Bolton 2004, Rubinstein et al 2007).

   b. **Indirect Methods**

   These were the anchor-based and distribution-based methods. The anchor-based method uses an external criterion, usually the PGIS. Anchor-based methods apply various relevant patient-rated variables, in this case changes in total BQ score (Rivicki et al 2008). A ROC curve was used to find the most accurate cut-off point that maximised sensitivity (true-positives) and specificity (true negatives) (Copay et al 2007, Houweling 2010; Maughan & Lewis 2010) in terms of categorising patients as

improved or not improved. A ROC curve was produced by plotting sensitivity (number of patients correctly identified as improved by the total change score on the BQ divided by the number of patients as improved according to the PGIS) against the specificity (number of patients correctly identified as unchanged/deteriorated by the total change score on the BQ divided by the number of unchanged/deteriorated patients according to the PGIS). For each possible cut off value the sensitivity (y-axis) was plotted against 1-specificity (x-axis) to generate the ROC curve (Maughan & Lewis 2010). Using the PGIS as the external criterion to categorise patients into improved (6/7)/not improved (1-5), the ROC curve was used to identify the cut-off change score in the total BQ score that demonstrated the 'minimal clinically important difference' [i.e. the value giving the best balance between sensitivity (improved patients) and specificity (non-improvers)] (Copay et al 2007; Maughan & Lewis 2010). This cut-off point was determined by calculating the Youden Index (sensitivity+specificity-1) which has been widely used to calculate the cut-off value in change scores that give the best balance between sensitivity and specificity (Fluss, Faraggi & Reiser 2005; Chen et al 2008; Maughan & Lewis 2010). The ROC curve method used in this study draws parallels with diagnostic studies (Revicki et al 2008).

The second indirect method, generally used to support the findings of the anchor-based method, was a distribution-based method (Revicki et al 2008; de Vet et al 2010). It has been found that 0.5 of the standard deviation of the group change score in an outcome measure corresponds to the minimally clinically important difference (Norman et al 2003; Copay et al 2007; Revicki et al 2008; Turner et al 2009; de Vet et al 2010). This method was used in the current study.

2. **Agreement between methods of categorising improvers**

Cohen's Kappa, a statistical measure of inter-rater agreement for categorical variables (e.g. improved/not improved) was used as the method of calculating agreement between the direct and the two indirect methods of calculating clinically significant improvement as described above. The Kappa statistic is generally thought to be more robust over simple percentage agreement as it takes into account agreement by chance (Puhan et al 2008).

STAGE 3

When assessing healthcare predictors, one often thinks of associations between these predictors such as someone who might smoke and drink being at greater risk of many common diseases such as heart disease, cancer and diabetes. Associations between the potential predictor variables at baseline were tested in this study using chi-squared analysis to gain understanding of their associations (Leboeuf-Yde et al 2004). The significance for the interaction was set at $p < 0.05$. All those variables that were significantly associated were entered into univariate analysis in the next stage (Stage 4) of the analysis. This was to ensure that even though predictors that were not associated with improvement by themselves, but were so when combined with another variable, were included in the final prediction model.

STAGE 4

Univariate logistic regression was used to calculate the crude odds ratios of all single potential predictors and those potential predictors in combination that were tested in Stage 3 for back/leg pain patients. This method was used for all three methods of assessing clinically significant improvement (i.e. direct, anchor-based and distribution-based methods). A choice had to be made at this stage to determine which method of determining clinically

significant improvement would be used to progress to the next stage of analysis. There is very little support for the direct method of assessing clinical significance in the literature. Almost without exception researchers over the last twenty or more years have agreed that the anchor-based method, looking at individual change rather than the distribution-based method which is based on a group cut-off, is the method of choice for assessing clinically significant change (Jordan et al 2006; Copay et al 2007; Revicki et al 2008; de Vet et al 2010; Terwee et al 2010). Moreover, the literature suggests the distribution-based methodology should be used as supportive evidence. For this reason the anchor-based methodology was used to progress to Stage 5 of the analysis.

STAGE 5

So as not to overlook a potential predictor of clinical improvement, all single potential predictors with a lax p-value of <0.1 and interactive potential predictors with a p-value of <0.5 from the univariate regression analysis were entered into multiple logistic regression model using the anchor-based method of assessing clinically significant improvement at treatment 4/5 and at treatment 10 (Henschke et al 2008). In addition, all patients who had a total BQ score of less than 13/70 at treatment 4/5 and a score of less than 19/70 at treatment 10 were eliminated from the analysis since change scores in these patients could not reach the threshold for categorisation as improved using the anchor-based method for this cohort.

STAGE 6

The final stage of analysis was to retain variables in the final multiple logistic regression model that were significant at p<0.05. The sensitivity and specificity of this model were calculated to determine the ability of the model to discriminate between improved and not-

improved patients. To calculate the diagnostic accuracy of the model, the area under the ROC curve (AUC) was calculated. An AUC of 0.5 indicates no discrimination, 0.7 to 0.8 acceptable discrimination and 0.9 excellent discrimination (Hosmer & Lemeshow 2000).

2.5.2 NECK, ARM AND/OR HEADACHE PATIENTS

The six stage analysis used for back/leg pain patients was calculated for patients whose presenting complaint was neck/shoulder/arm pain, with or without headaches. The methodology used was exactly the same as that used for back/leg pain patients except that all patients with an initial BQ score of less than 14.5/50 were excluded. The reason for this was that a change score in excess of this figure was required to demonstrate the patient had improved using the anchor-based methodology for this cohort of patients at Stage 2 of data analysis.

2.6 ETHICAL APPROVAL

Approval for this study was obtained from the Ethics Sub-Committee of the Anglo-European College of Chiropractic (see appendix V).

Chapter 3: Results

The throughput of patients enrolled in this study is given in Figure 2.

```
┌─────────────────────────────────────┐
│          4724 New Patients          │
└─────────────────────────────────────┘
                    │
      ┌─────────────────────────────┐
      │  818 babies/children        │
      │  under 16 years old         │
      │                             │
      │  51 frail/elderly or        │
      │  not fluent in English      │◄──┤
      │                             │
      │  72 did not complete        │
      │  BQ (did not feel it        │
      │  was an appropriate         │
      │  outcome measure)           │
      └─────────────────────────────┘
                    │
┌─────────────────────────────────────┐
│    3783 completed pre-treatment BQ  │
└─────────────────────────────────────┘
                    │
┌────────────────────────────────────┐
│  79 patients without back or neck  │
│  pain as their primary complaint   │
│                                    │
│  1047 patients did not attend for 4│
│  treatments or the receptionist    │
│  forgot to ask them to complete    │◄──┤
│  the post-treatment BQ or patient  │
│  said they would complete at       │
│  following appointment and did not.│
└────────────────────────────────────┘
                    │
        ┌───────────────────────────┐
        │  2657 completed follow-up │
        │  BQ at visit 4/5          │
        └───────────────────────────┘
                    │
        ┌───────────────────────────┐
        │  467 patients followed-up │
        │  at visit 10              │
        └───────────────────────────┘
```
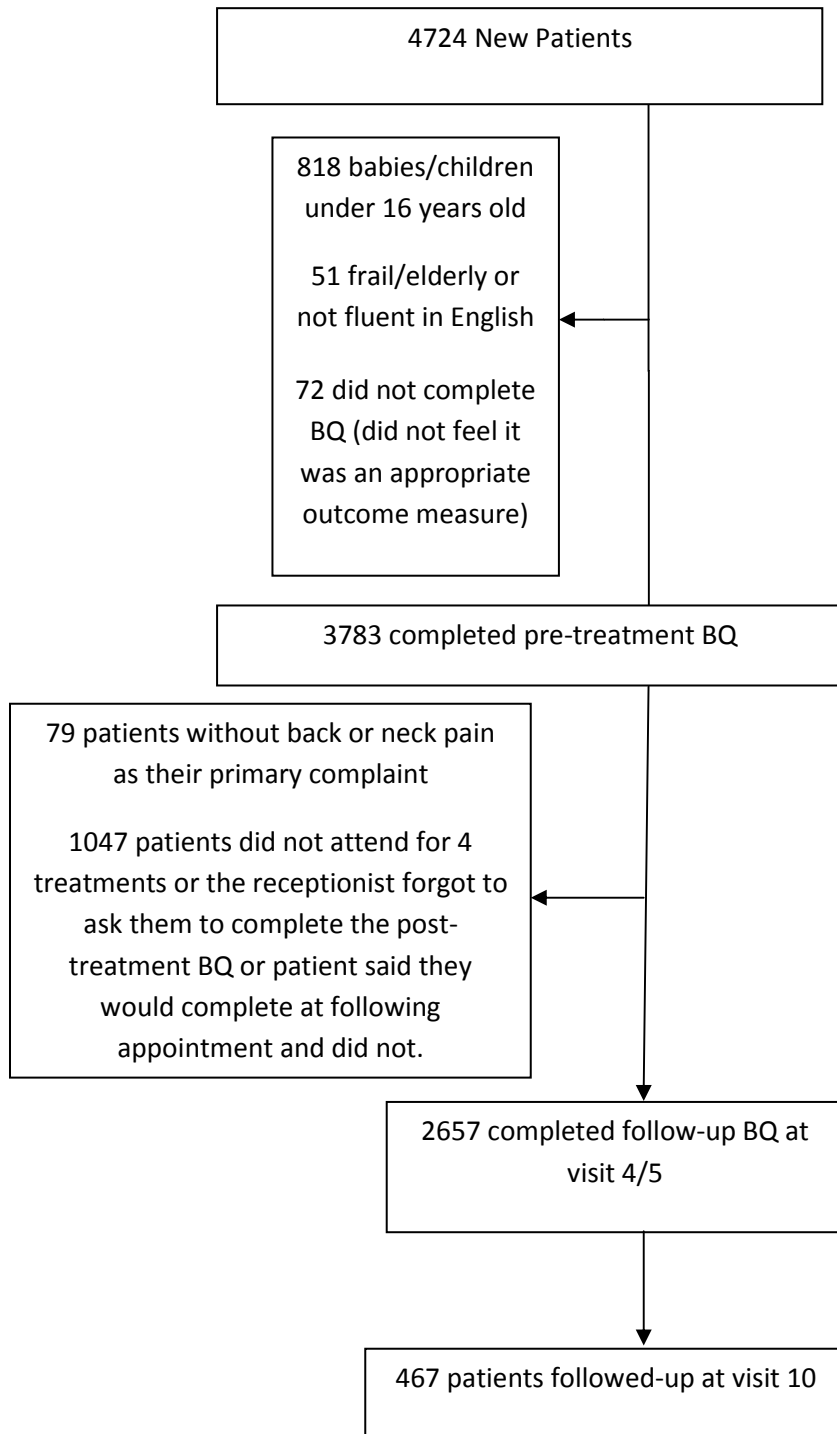
Figure 2. Flow chart of patients in study and reasons for loss at follow-up

The total number of new patients that presented at the clinic was 4724, of which 941 patients (nearly twenty percent) had to be discounted as they did not complete the pre-treatment BQ for reasons shown in Figure 2. Almost twenty-eight percent of eligible patients did not complete the first post-treatment BQ at visit 4/5 either because they discontinued care before their 4th treatment, the receptionist forgot to ask them to fill in the questionnaire or they said they would do so at the follow-up visit and for one reason or another, this did not happen. As there was no follow-up questionnaire for these patients, their initial BQ was not included in the dataset.

The total number of patients in the dataset for the purposes of this study was therefore 2657 (fifty-six percent of all new patients and seventy percent of the new patients who completed the pre-treatment BQ) of whom 1871 had back and/or leg pain and 1267 had neck/shoulder arm pain, with or without headaches (see Figure 2). These were not mutually exclusive and 481 of the patients had both back and/or leg pain as well as neck and/or arm pain with or without headaches as their presenting complaint. The data were collected between November 2001 and January 2010.

ANALYSIS OF BACK/LEG PAIN PATIENTS

3.1.1 STAGE 1: DESCRIPTIVE STATISTICS OF STUDY POPULATION

Of the 2657 patients enrolled in the study, 1871 had back and/or leg pain. Table 2 gives the baseline data for these patients. The gender distribution was almost equal (male=fifty one percent) and the mean age of the patients was 41±14.5 years. Interestingly more than half the study population had radiating leg pain. Moreover, more than two-thirds (sixty nine percent) of patients had experienced a similar complaint in the past and more than half of patients (fifty seven percent) were sub-acute or chronic having had the current episode for

more than 7 weeks. Despite all being private patients, it is interesting to note that only

around two-thirds (sixty six percent) expected to recover or improve with treatment.

Table 2. Description of low back/leg pain patients at baseline (n=1871)

Age	41.0 years (SD±14.5)
Gender (Male)	50.9% (n=952)
Radiating Pain	53.1% (n=924)
Widespread Pain	5.7% (n=104)
Trauma	25.6% (n=474)
Similar complaint in past	68.8% (n=1261)
Current episode duration <7 weeks	57.4% (n=1061)
Medication use (lot of the time)	24.5% (n=453)
Patient expectation-recover/improve	65.8% (n=1225)
-stay the same	29.4% (n=547)
-get worse	4.8% (n=90)
Work status -employed	79.1% (n=1460)
-retired	9.9% (n=183)
-seeking work	2.3% (n=42)
-working in home	6.3% (n=116)
-student	2.4% (n=44)
Job satisfaction (satisfied with work status)	91.6% (n=1603)
Work expectation (work normally in 6 months)	92.6% (n=1596)
Current/previous smoker	53.7% (n=1003)
Current alcohol drinker	75% (n=1399)
Level of physical activity-more/same as peers	80.9% (n=1506)
-less than peers	19.1% (n=355)
General well-being –excellent/good	86.4% (n=1610)
-fair/poor	13.6% (n=254)
BQ pain intensity (maximum 10/10)	5.81(SD±2.17)
BQ disability (Activities of Daily Living) (maximum 10/10)	5.33 (SD±2.74)
BQ interference in social life (maximum 10/10)	4.91 (SD±3.09)
BQ anxiety (maximum 10/10)	4.81 (SD±2.84)
BQ depression (maximum 10/10)	3.24 (SD±2.92)
BQ work fear avoidance (maximum 10/10)	4.65 (SD±3.06)
BQ locus of control (maximum 10/10)	4.92 (SD±2.64)
BQ total score (maximum 70/70)	33.4 (SD±14.1)
Pain diagram (number of areas shaded –maximum 52/52)	7.66 (SD±6.3)
Pain diagram inappropriate signs	12.9% (n=242)

Outcomes at 4/5th and 10th Treatment

The mean duration between the initial consultation and the 4/5th treatment was 16.6 ±9.5 days and at the 10th treatment the mean duration was 50.7±14.0 days. At the time of treatment 4/5 (n=1871), the mean total BQ score had reduced from 33.4 (SD±14.1) to 17.2 (SD±12.8). By treatment 10 (n=289), the mean total BQ score had reduced to 12.7 (SD±11.8) (Figure 3).
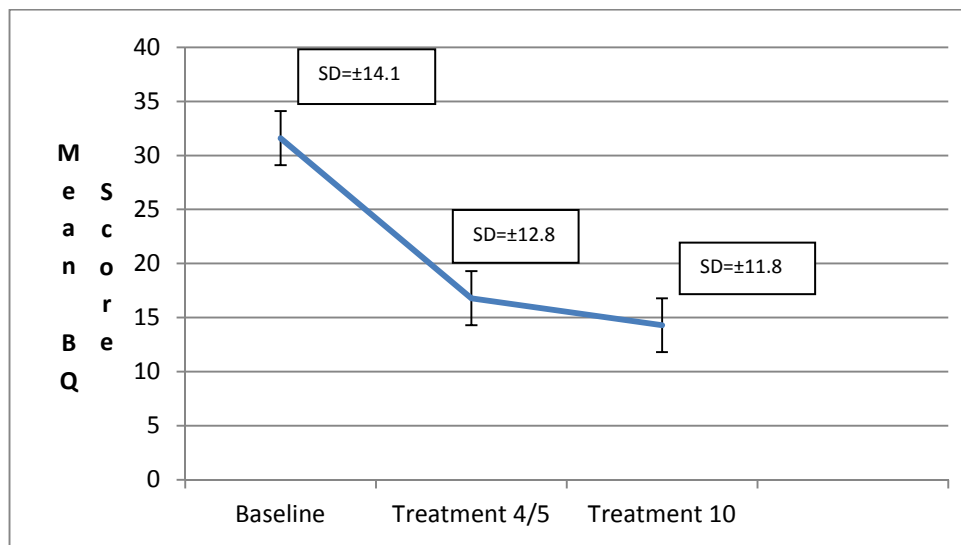
Figure 3. Change in baseline total BQ scores in low back/leg pain patients undergoing treatment

Of those patients in paid employment, most patients had not taken any sick leave, and of those who had, just over four percent at treatment 4/5 and one-and-a-half percent at treatment 10 were still on sick leave (Table 3). Almost all the patients were satisfied with the treatment they received, and felt that the treatment had met their expectations.

Table 3. Percentages of patient work status and treatment satisfaction

	Treatment 4/5	Treatment 10
In paid employment	77.8%	75.3%
Not in paid employment	22.2%	23.7%
No sick leave	61.8%	70.1%
Return to work after sick leave	9.9%	6%
Still on sick leave	4.3%	1.4%
Treatment satisfaction	99.8%	100%
Treatment expectation met	96.4%	95.5%

3.1.2 STAGE 2: CATEGORISING IMPROVEMENT IN PATIENTS

Table 4 shows improvement in patients at treatments 4/5 and 10 using the direct and indirect methods of categorisation.

Table 4. Proportion of patients improved using three methods of categorising improvement

		Improved
Patient Global Improvement Scale (Direct Method)	**Treatment 4/5**	**59.1% (n=1102)**
	Treatment 10	73% (n=211)
Anchor-based method (Indirect Method)	**Treatment 4/5**	**58.1% (n=1082)**
	Treatment 10	54.6% (n=159)
Distribution Method (Indirect Method)	**Treatment 4/5**	**72.1% (n=1343)**
	Treatment 10	79% (n=230)

The first method of classifying patients as improved/not improved was the 'direct' method using the PGIS with fifty nine percent of patients improved at treatment 4/5 and seventy-three percent at treatment 10 (Table 4). The second method of classifying improved/not improved patients was the anchor-based method. It was found that fifty eight percent of

patients were improved at treatment 4/5 and fifty five percent at treatment 10. A ROC curve was used to find the best cut-off in the total BQ change score between improved and not-improved patients. The cut-off with the best balance between sensitivity and specificity (Youden Index) was a change score of 13/70 at treatment 4/5 and 19/70 at treatment 10. The third method of classifying patients as improved/not improved was the distribution-based method which demonstrated seventy two percent of patients were improved at the treatment 4/5 and seventy nine percent at the treatment 10. These values amounted to a reduction of at least 7 points in total BQ score at the treatment 4/5 and 8 point reduction at treatment 10. The change score of 7 and 8 points was half the standard deviation of the group change score at treatment 4/5 and treatment 10 respectively.

Table 5 demonstrates the level of agreement between the direct, anchor and distribution-based methods of classifying patients as improved. The strongest agreement was found between the anchor and distribution-based methods of calculating clinically significant improvement with moderate agreement between the PGIS and anchor-based method, and the weakest agreement was between the PGIS and distribution-based method.

Table 5. Agreement between three methods of assessing clinically significant improvement

	Methods Compared	Kappa Statistic	95% Confidence Interval
Treatment 4/5	*PGIS/Anchor-based	0.44	0.39-0.48
	*PGIS/Distribution	0.39	0.35-0.44
	Anchor/Distribution	0.70	0.66-0.73
Treatment 10	*PGIS/Anchor-based	0.43	0.33-0.54
	*PGIS/Distribution	0.38	0.24-0.51
	Anchor/Distribution	0.48	0.38-0.59

*= Patient Global Improvement Scale

3.1.3 STAGE 3: ASSOCIATIONS BETWEEN BASELINE VARIABLES

Before investigating possible significance of interactions between potential predictor variables at baseline and improvement at treatment 4/5 and treatment 10, analysis was carried out to look at the associations between potential predictors at baseline. Table 6 demonstrates the associations between the baseline variables using Chi-Squared analysis with the level of statistical significance set at p=0.05. Cells left blank on Table 6 indicated no statistically significant association between baseline variables.

As can be seen in Table 6 there were a significant number of statistically significant associations between variables at baseline. All associations that were significant were entered into the subsequent univariate logistic regression analysis, in addition to all individual baseline variables themselves.

Table 6. Associations between potential predictor variables at baseline (*=chi-squared analysis p<0.5)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Trauma (1)	-																		
Treatment Expectation (2)	*<0.001	-																	
Pain diagram (3)		*<0.001	-																
Gender (4)			*<0.001	-															
Employment status (5)	*0.01				-														
Age (6)			*<0.001			-													
Inappropriate signs (7)	*0.03	*<0.001	*<0.001	*<0.001			-												
Radiating pain (8)			*<0.001			*<0.001	*<0.001	-											
Widespread pain (9)	*0.02		*<0.001	*<0.001	*0.03		*<0.001	*0.03	-										
Similar complaint (10)	*<0.001		*0.05	*<0.001						-									
Pain Duration (11)		*<0.001	*<0.001		*<0.001	*0.04	*<0.001		*<0.001		-								
Medication use (12)	*<0.001		*0.02	*<0.001		*<0.001	*<0.001	*<0.001	*<0.001			-							
Work satisfaction (13)		*0.01	*0.03		*<0.001		*0.01	*0.02			*<0.001		-						
Work Expectation (14)		*<0.001			*<0.001	*<0.001	*<0.001		*0.02		*<0.001		*<0.001	-					
Smoker (15)				*<0.001				*0.01				*<0.001	*0.02		-				
Alcohol (16)			*0.01	*<0.001		*<0.001	*<0.001		*<0.001		*<0.001	*0.01	*0.03	*0.02	*<0.001	-			
Physical activity (17)		*<0.001	*<0.001	*<0.001		*0.03	*<0.001	*0.03	*<0.001	*0.04	*<0.001	*<0.001	*0.02	*<0.001		*<0.001	-		
General health (18)		*<0.001	*<0.001		*<0.001		*<0.001		*<0.001	*0.01	*<0.001	*0.01	*<0.001	*<0.001	*0.01	*<0.001	*<0.001	-	
Severity (BQ<30/70) (19)		*<0.001	*<0.001	*<0.001		*0.05	*<0.001	*<0.001	*0.01		*<0.001	*<0.001	*<0.001			*<0.001	*<0.001	*<0.001	-

Potential Predictor in column 1 given numbers which identify potential predictors in row 1

3.1.4 STAGE 4: UNIVARIATE REGRESSION ANALYSIS OF POTENTIAL PREDICTORS

Univariate Regression Analysis was carried out for the individual baseline variables (potential predictors) and significant combined variables from Stage 3. These results can be found in Table A in appendix vi, for each of the direct, anchor-based and distribution-based methods for categorising improved patients. For all subsequent analysis, the anchor-based method of improvement was used. Table 7 shows those single predictors that were significantly ($p<0.1$) associated with improvement at treatment 4/5 using the anchor-based method.

Table 7. Baseline variables significantly associated with improvement at treatment 4/5

Variable	Odds Ratio	95% Confidence Interval	P value (significance<0.1)
Chronicity (<7weeks)	2.67	2.18 to 3.27	<0.01
Medication use	1.23	0.98 to 1.54	0.08
Work Expectation	1.75	1.20 to 2.55	<0.01
Smoker	1.18	0.97 to 1.44	0.09
General Health	1.48	1.13 to 1.94	0.01
Pain Diagram (<8/52)	1.28	1.05 to 1.56	0.02
Treatment Expectation	2.10	1.71 to 2.58	<0.01
Age (<50)	1.27	1.02 to 1.59	0.04
Baseline BQ (<30/70)	0.20	0.16 to 0.25	<0.01

Those interactions between baseline variables that were significant ($p<0.05$) from Stage 3 of the analysis were similarly entered into a univariate regression analysis with improvement at treatment 4/5 using the anchor-based method as the dependent variable. Those interactions that were significantly ($p<0.05$) associated with improvement are shown in Table 8. The significance level was set at $p<0.05$ because of the number of significant variables identified.

Table 8. Baseline interaction variables associated with improvement at treatment 4/5

Interactive Variables	Odds Ratio	95% Confidence Interval	P value (significance <0.05)
TreatmentExpectationXTrauma	1.94	1.56 to 2.41	<0.01
TreatmentExpectationXPainDiagram	1.80	1.39 to 2.32	<0.01
TreatmentExpectationXInappropriateSigns	1.94	1.56 to 2.41	<0.01
Chronicity<7weeksXTreatmentExpectation	2.60	2.05 to 3.29	<0.01
TreatmentExpectationXJobsatisfaction	2.10	1.26 to 3.51	0.01
TreatmentExpectactionXWorkNormally	2.27	1.35 to 3.86	<0.01
TreatmentExpectationXGeneralHealthPerception	2.01	1.36 to 2.97	<0.01
AgeXPainDiagram	1.79	1.27 to 2.52	<0.01
PainDiagramXRadiating Pain	1.40	1.06 to 1.85	0.02
WidespreadPainXPainDiagram	1.27	1.04 to 1.56	0.02
Chronicity<7weeksXPainDiagram	1.86	1.47 to 2.36	<0.01
PainDiagramXMedicationUse	1.29	1.04 to 1.59	0.02
PainDiagramXPhysicalActive	1.19	1.03 to 1.37	0.02
PainDiagramXGeneralHealthPerception	1.81	1.26 to 2.61	<0.01
BQ<30/70XPain Diagram	0.56	0.44 to 0.67	<0.01
BQ<30/70XGender	0.47	0.34 to 0.53	<0.01
Chronicity<7weeksXAge	2.08	1.54 to 2.83	<0.01
BQ<30/70XAge	0.58	0.43 to 0.78	<0.01
Chronicity<7weeksXInappropriateSigns	2.50	2.03 to 3.08	<0.01
GeneralHealthPerceptionXInappropriate Signs	1.40	1.03 to 1.92	0.03
BQ<30/70XInappropriate Signs	0.26	0.21 to 0.32	<0.01
BQ<30/70XRadiating Pain	0.36	0.28 to 0.47	<0.01
Chronicity<7weeksXWidespreadPain	5.56	2.08 to 3.15	<0.01
WidespreadPainX WorkNormally	1.82	1.21 to 2.74	<0.01
WidespreadPainXGeneralHealthPerception	1.59	1.18 to 2.13	<0.01
BQ<30/70XWidespreadPain	0.24	0.19 to 0.3	<0.01
Chronicity<7weeksXJobSatisfaction	1.59	1.02 to 2.48	0.04
Chronicity<7weeksXWorkNormally	2.08	1.27 to 3.4	<0.01
Chronicity<7weeksXPhysicalActivity	1.63	1.21 to 2.2	<0.01
Chronicity<30/70XGeneralHealthPerception	1.91	1.35 to 2.71	<0.01
MedicationUseXSmoking	1.30	1.06 to 1.59	0.01
GeneralHealthPerceptionXMedicationUse	1.65	1.19 to 2.27	<0.01
BQ<30/70XMedicationUse	0.32	0.26 to 0.4	<0.01
Job satisfactionXSmoking	1.77	1.01 to 3.2	0.05
GeneralHealthPerceptionXJobSatisfaction	2.14	1.18 to 3.89	0.01
BQ<30/70XAlcohol	0.50	0.38 to 0.66	<0.01
GeneralHealthPerceptionXPhysicalActivity	1.70	1.17 to 2.47	0.01
BQ<30/70XPhysicalActivity	0.72	0.54 to 0.97	0.03

A similar analysis was carried out for improvement using the anchor-based method at

treatment 10. Table 9 shows the baseline variables significantly associated with

improvement at treatment 10 and Table 10 shows the significant interactive baseline

variables.

Table 9. Baseline variables significantly associated with improvement at treatment 10

Variable	Odds Ratio	95% Confidence Interval	P value (significance <0.1)
Chronicity<7weeks	4.08	2.34 to 7.13	<0.01
Pain diagram<8/52	1.98	1.17 to 3.34	0.01
Treatment Expectation	3.68	2.14 to 6.35	<0.01
Baseline BQ score<30/70	0.23	0.13 to 0.40	<0.01

Table 10. Baseline interaction variables significantly associated with improvement at treatment 10

Interactive Variables	Odds Ratio	95% Confidence Interval	P value (significance<0.05)
TreatmentExpectationXTrauma	3.19	1.82 to 5.6	<0.01
TreatmentExpectationXPainDiagram	2.61	1.39 to 4.89	<0.01
TreatmentExpectationXInappropriateSigns	2.80	1.61 to 4.87	<0.01
ChronicityXTreatmentExpectation	3.96	2.19 to 7.19	<0.01
TreatmentExpectationXPhysicalActivity	2.43	1.02 to 5.8	0.05
Pain DiagramXInappropriateSigns	1.79	1.06 to 3.04	0.03
Pain diagramXRadiatingPain	2.12	1.11 to 4.04	0.02
WidespreadPainXPainDiagram	2.16	1.28 to 3.66	<0.01
ChronicityXPain Diagram	2.16	1.21 to 3.83	0.01
Pain DiagramXMedicationUse	2.22	1.27 to 3.86	0.01
Widespread PainXInappropriateSigns	0.47	0.24 to 0.92	0.03
Chronicity<7weeksXInappropriate Signs	2.88	1.70 to 4.91	<0.01
AlcoholXInappropriateSigns	0.37	0.18 to 0.77	0.01
Baseline BQ<30/70XInappropriateSigns	.20	0.12 to 0.35	<0.01
Chronicit<7weeksXWidespreadpain	3.62	2.11 to 6.24	<0.01
Chronicity<7weeksXPhysicalActivity	2.16	1.02 to 4.57	0.04
Baseline BQ<30/70XMedicationUse	0.43	0.25 to 0.74	<0.01

3.1.5 STAGE 5: MULTIVARIABLE LOGISTIC REGRESSION ANALYSIS FOR OUTCOME PREDICTORS

This stage of analysis involved entering the significant baseline variables from the univariate regression model in Stage 4 into a multivariable regression analysis in order to determine those variables that were independently associated with outcome when adjusting for all other variables in the model. This was done in three steps: first the single significant predictor variables, second the interactive significant variables and thirdly, those variables that remained significant ($p<0.05$) in the multivariable models from the first two steps were entered together into the final predictive multivariable model.

I. PREDICTORS OF OUTCOME AT TREATMENT 4/5

a. STEP 1-Single Predictors

The 9 single predictors from Stage 4 (Table 7) at a significance level of $p<0.1$ were entered into a multiple logistic regression model for improvement at treatment 4/5. Five single predictors remained significant in the multiple regression model (Table 11). Without the predictor model, sixty three and a half percent of the patients were correctly classified as improved. Using this predictor model, seventy two percent of the patients were correctly classified as reliably improved.

Table 11. Baseline variables independently associated with improvement at treatment 4/5

Variable	Odds Ratio	95% Confidence Interval	P value (significance p<0.05)
Chronicity (<7 weeks)	2.09	1.64 to 2.68	<0.01
General Health Perception	1.59	1.15 to 2.2	0.01
Pain Diagram (<8/52)	1.32	1.04 to 1.67	0.02
Treatment Expectation	1.38	1.07 to 1.78	0.01
Baseline BQ score(<30/70)	0.20	0.16 to 0.25	<0.01

b. STEP 2-Interactive Predictors

The 38 interactive predictors from Stage 4 (Table 8) were entered into a multivariable logistic regression model. Only 2 of the interactive predictors remained significant in this model (Table 12). The percentage of patients correctly classified as improved was sixty three percent prior to the modelling and seventy two percent were correctly classified as improved in this model.

Table 12. Baseline interactive variables independently associated with improvement at treatment 4/5

Interactive Variables	Odds Ratio	95% Confidence Interval	P value (significance <0.05)
Baseline BQ<30/70Xleg pain	0.66	0.44 to 0.98	0.04
Baseline BQ<30/70Xwidespread pain	0.42	0.21 to 0.82	0.01

c. STEP 3-Final Model

The final model in stage 5 included all significant single (Table 11) and interactive (Table 12) predictors. All interactive predictors dropped out of the final model but 5 single predictors remained significant (Table 13). This final model increased the predicted percentage of patients who were improved at treatment 4/5 from sixty three percent to seventy two percent.

Table 13. Final model showing all variables independently associated with improvement at treatment 4/5

Variable	Odds Ratio	95% Confidence Interval	P value (significance p<0.05)
Chronicity <7 weeks	2.26	1.77 to 2.88	<0.01
General Health Perception	1.52	1.10 to 2.1	0.01
Pain Diagram (<8/52)	1.30	1.03 to 1.65	0.03
Treatment Expectation	1.46	1.14 to 1.87	<0.01
Baseline BQ Score(<30/70)	0.13	0.07 to 0.25	<0.01

II. PREDICTORS OF OUTCOME AT TREATMENT 10

a. STEP 1-Single Predictors

The 4 single predictors from Stage 4 (Table 9) at a significance level of p<0.1 were entered into a multiple logistic regression model for improvement at treatment 10. All 4 single predictors remained significant in the multiple regression model although the confidence intervals were wider in the results for treatment 10 as the number of subjects in this model was 248 compared to 1538 patients at treatment 4/5 (Table 14). Without the predictor model, sixty three percent of the patients were correctly classified as improved. Using this

predictor model, the number of patients correctly classified as improved increased to seventy two percent.

Table 14. Baseline variables independently associated with improvement at treatment 10

Variable	Odds Ratio	95% Confidence Interval	P-value (significance <0.05)
Chronicity (<7 weeks)	2.83	1.53 to 5.22	<0.01
Pain Diagram (<8/52)	1.91	1.05 to 3.49	.04
Treatment Expectation	2.77	1.51 to 5.1	<0.01
Baseline BQ score(<30/70)	0.23	0.13 to 0.44	<0.01

b. STEP 2-Interactive Predictors

The 17 interactive predictors from stage 4 (Table 10) at a significance level of p<0.05 were entered into a multivariable logistic regression model. Five interactive predictors remained significant in this model (Table 15). The percentage of patients improved was sixty three percent prior to the modelling and seventy eight percent were correctly classified as improved using this model. For the same reasons stated in Step 1 there were wider confidence intervals.

Table 15. Baseline interactive variables independently associated with improvement at treatment 10

Interactive Variables	Odds Ratio	95% Confidence Interval	P value (significance <0.05)
TreatmentExpectationXTrauma	5.22	1.43 to 19.12	0.01
PainDiagramXInappropriate signs	17.23	2.51 to 118.14	0.01
Chronicity<7weeksXPainDiagram	0.23	0.03 to 0.52	0.01
AlcoholXInappropriate signs	0.23	0.08 to 0.62	<0.01
BQ score<30/70XInapproriate signs	0.06	0.01 to 0.23	<0.01

c. STEP 3-Final Model

The final model in stage 5 included all significant single (Table 14) and interactive (Table 15) baseline predictors. Two of the single predictors and four of the interactive predictors remained in the model for patients at treatment 10, albeit with wider confidence intervals than the final model at treatment 4/5 (Table 16). This model increased the predicted percentage of patients who were improved at treatment 10 from sixty four percent to eighty one percent.

Table 16. Final model showing all variables independently associated with improvement at treatment 10

Variables	Odds Ratio	95% Confidence Interval	P value (significance <0.05)
Chronicity (<7 weeks)	12.67	4.22 to 38.05	<0.01
Baseline BQ Score (<30/70)	22.45	1.49 to 37.44	0.02
PainDiagramXInappropriate Signs	7.66	1.61 to 36.53	0.01
ChronicityXPainDiagram	0.12	0.03 to 0.47	<0.01
AlcoholXInappropriate Signs	0.51	0.33 to 0.79	<0.01
Baseline BQ<30/70XInapproriate Signs	0.08	0.19-to 0.35	<0.01

In conclusion, at treatment 4/5 and at treatment 10, the single predictors of chronicity (<7 weeks duration) and a baseline BQ score of less than 30/70 were significant in the final multivariable model although the odds ratio for the BQ score was below 1 which meant that a total BQ score in excess of 30/70 was required as a predictor of improvement. The way the patient rated their general health, their treatment expectation and the pain diagram were significant predictors at treatment 4/5. At treatment 10 there were four interactive predictors, inappropriate markings on the pain diagram and the area shaded on the pain diagram itself, inappropriate markings on the pain diagram with regular alcohol consumption, inappropriate markings on the pain diagram with the baseline BQ score and chronicity (<7 weeks duration) with the pain diagram were associated with improvement.

3.1.6 STAGE 6: PROPERTIES OF FINAL PREDICTION MODEL

i.   *Treatment 4/5*

Only those predictors significantly associated with improvement at treatment 4/5 (Table 13) were entered into the final model. This final stage of analysis calculated the properties of the final predictive models. These were the sensitivity (prediction of patients who were improved and were correctly identified) at eighty six percent, specificity (those who did not improve who were correctly identified) at forty nine percent, the positive predictive value of the model (the chance an improved patient has been correctly classified) at seventy four percent and the negative predictive value (the chance that a patient who has not improved has been correctly classified) at sixty seven percent. The area under the ROC curve gave the accuracy of the model in distinguishing between patients improved and not improved as categorised by the anchor-based method (Table 17) and was 0.75 demonstrating reasonably good discriminative ability.

Table 17. Final predictive model of improvement at treatment 4/5 in back and/or leg pain patients

	Variables Tested	Odds Ratio and 95% Confidence Interval	1. Sensitivity (95% Confidence Interval) 2. Specificity (95% Confidence Interval) 3. Numbers correctly classified improved (95% Confidence Interval) 4. Number correctly classified not improved (95% Confidence Interval) 5. Area under the ROC (95% Confidence Interval)
	Chronicity (<7 weeks)	2.20 (1.74 to 2.78)	
	General Health Perception	1.50 (1.1 to 2.04)	
	Pain diagram (<8/52)	1.35 (1.08 to 1.69)	
	Treatment Expectation	1.41 (1.11 to 1.75)	
	Baseline BQ (<30/70)	0.19 (0.15 to 0.24)	
FINAL MODEL			1. 86.1% (83.9 to 88%) 2. 48.5% (44.6 to 52.5%) 3. 74.1% (71.6 to 76.5%) 4. 67.0% (62.6 to 71.2%) 5. 0.75 (0.73 to 0.78)

ii. *Treatment 10*

Only those predictors significantly associated with improvement at treatment 10 (Table 15) were entered into the final model. This final stage of analysis calculated the properties of the final predictive model. These were the sensitivity ninety two percent (ability of this model to predict which patients would recover), specificity forty nine percent (ability of the model to predict which patients would not improve), the positive predictive value was seventy six percent (ability of the model to correctly categorise which patients had improved), negative predictive value of seventy nine percent (ability of the model to correctly categorise which patients would not improve) and area under the ROC curve or the ability of the model to distinguish between who was improved/not improved and this was good at 0.83 (Table 18).

The final predictive model for back/leg pain patients at treatment 10 included the two single

predictors shown to be statistically significant as well as the four statistically significant

interactive predictors although one of the single predictors, baseline BQ score, dropped out

of the final model.

Table 18. Final predictive model of improvement at treatment 10 for back and/or leg pain

	Variables Tested	Odds Ratio and 95% Confidence Interval	1. Sensitivity (95% CI) 2. Specificity (95% CI) 3. Numbers correctly classified improved (95% CI) 4. Number correctly classified not improved (95% CI) 5. Area under the ROC (95% CI)
	Chronicity (<7 weeks)	11.63 (4.62 to 29.27)	
	PainDiagramXInappropriate signs	8.17 (3.41 to 19.59)	
	Pain DiagramXChronicity (<7 weeks)	0.17 (0.06 to 0.49)	
	AlcoholXInappropriate Signs	0.26 (0.11 to 0.61)	
	Baseline BQ<30/70XInappropriate Signs	0.15 (0.08 to 0.31)	
FINAL MODEL			1. 92.4% (87.1 to 95.6%) 2. 48.9% (38.8 to 59.0%) 3. 75.9% (69.4 to 81.4%) 4. 78.6% (66.2 to 87.3%) 5. 0.83 (0.77 to 0.88)

ANALYSIS OF NECK/ARM PAIN (with or without headache) PATIENTS

3.2.1 STAGE 1: DESCRIPTIVE STATISTICS OF STUDY POPULATION

Table 19 gives the baseline data for 1267 patients with neck/arm with or without headache. The mean age of the patients was almost the same as the back pain group but in common with the literature, there were more females in the neck pain group (fifty eight percent). As with the back pain group, around two-thirds of this group has experienced a similar problem in the past (sixty three percent) and nearly half the group (forty seven percent) were subacute or chronic patients (duration of current episode more than 7 weeks). Despite being private patients, forty percent of this group expected their condition to stay the same or get worse. The total pre-treatment BQ score was lower than that of the back pain group (31.6 for neck pain versus 33.4 for back pain) but there was slightly larger areas shaded on the pain diagram (8.4 squares for neck pain versus 7.66 for back pain) and the patients were more likely to make inappropriate markings in the pain diagram (twenty percent of neck pain versus thirteen percent for back pain) in the neck pain group. Seventeen percent of this patient group had experienced headaches with their neck/shoulder arm pain.

Table 19. Description of neck/arm (with or without headache) patients at baseline (n=1267)

Age	40.1(SD±14.4)
Gender (Male)	42.3% (n=536)
Radiating Pain	43.1% (n=509)
Widespread Pain	7.4% (n=91)
Trauma	26.1% (n=325)
Similar Complaint in past	63.3% (n=785)
Current episode duration <7 weeks	47.3% (n=593)
Medication use (lot of the time)	25.3% (n=316)
Patient expectation-recover/improve	60.5% (n=759)
-stay the same	33.6% (n=420)
-get worse	5.8% (n=73)
Work status -employed	79.1% (n=986)
-retired	8% (n=100)
-seeking work	3.4% (n=43)
-working in home	6.3% (n=78)
-student	3.2% (n=40)
Job satisfaction (satisfied with status)	89% (n=1055)
Work expectation (work normally in 6 months)	93.6% (n=1091)
Current/previous smoker	50.9% (n=644)
Current alcohol drinker	69.5% (n=876)
Level of physical activity-more/same as peers	79.7% (n=1006)
-less than peers	20.3% (n=256)
General well-being –excellent/good	84.3% (n=1065)
-fair/poor	15.5% (n=196)
BQ subjective pain (max10/10)	5.58(SD±2.15)
BQ disability (max 10/10)	4.67 (SD±2.74)
BQ interference in social life (max10/10)	4.15 (SD±3.05)
BQ anxiety (max10/10)	4.96 (SD±2.83)
BQ depression (max 10/10)	3.26 (SD±2.92)
BQ work fear avoidance (max 10/10)	4.43 (SD±3.05)
BQ locus of control (max 10/10)	4.79 (SD±2.72)
BQ total score (max 70/70)	31.6 (SD±14.2)
Pain diagram (no. of areas shaded–max 52/52)	8.41 (SD±7.3)
Pain diagram inappropriate signs	19.9% (n=251)

Outcomes at 4/5th and 10th Treatment

The mean duration between the initial consultation and the 4/5[th]treatment as 17±11.5 days,

and at the 10th treatment the mean duration was 52.2±16 days. At the time of treatment 4/5

(n=1246) the mean total BQ score had reduced from 31.6 (SD±14.2) to 16.8 (SD±12.1). At

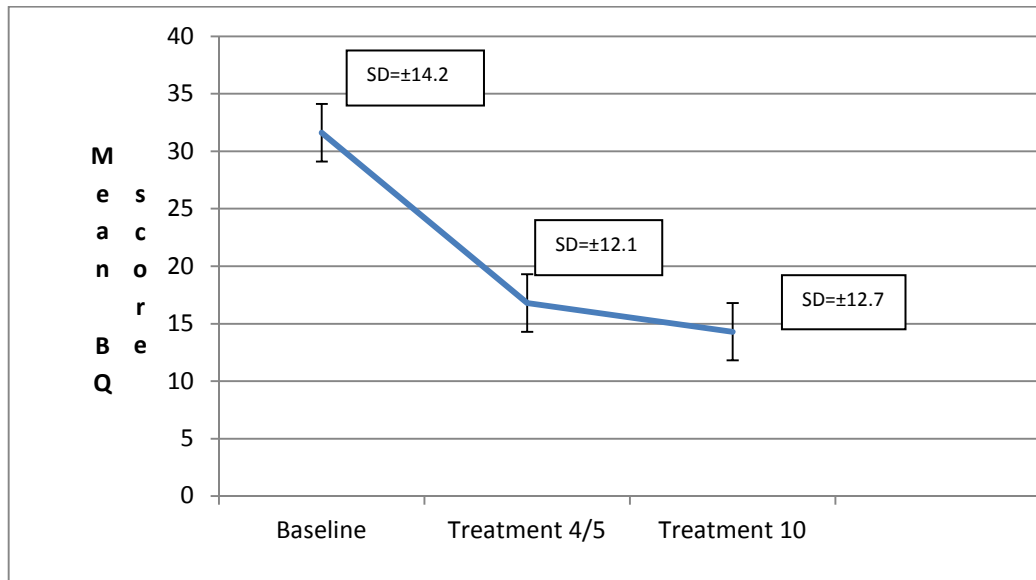treatment 10 (n=160) the mean total BQ score had reduced to 14.3 (SD±12.7) (Figure 4).

Figure 4. Change in baseline total BQ scores in neck/arm pain patients undergoing treatment

Of those in paid employment, most of the patients had not taken any sick leave, and of those who had just under four percent were still on sick leave at treatment 4/5 and just over one percent were still off work at treatment 10. Almost all of the patients were satisfied with the treatment they had received, and felt the treatment had met their expectations (Table 20)

Table 20. Percentages of patient work status and treatment satisfaction

	Treatment 4/5	Treatment 10
In paid employment	71.9%	75.6%
Not in paid employment	28.1%	24.4%
No sick leave	64%	72.1%
Return to work after sick leave	7.9%	3.5%
Still on sick leave	3.8%	1.2%
Treatment satisfaction	99.8%	100%
Treatment expectation met	96.6%	95.3%

3.2.2 STAGE 2: CATEGORISING IMPROVEMENT IN PATIENTS

The first method of classifying patients as improved/not improved was the 'direct' method using the PGIS with fifty six percent of patients improved at the treatment 4/5 and sixty eight percent at the treatment 10 (Table 21).

Table 21. Proportion of neck/arm (with or without headache) patients improved at treatment 4/5 and 10 using direct and indirect methods

		Improved
Patient Global Improvement Scale (Direct Method)	**Treatment 4/5**	**56.2% (n=712)**
	Treatment 10	68% (n=119)
Anchor-based method (Indirect Method)	**Treatment 4/5**	**47.3% (n=598)**
	Treatment 10	58.5% (n=103)
Distribution Method (Indirect Method)	**Treatment 4/5**	**68% (n=860)**
	Treatment 10	73.3% (n=129)

The second method of classifying improved/not improved patients was the anchor-based method. It was found that forty seven percent of patients were improved at treatment 4/5 and fifty nine percent at treatment 10. A ROC curve was used to find the best cut-off in the change in the total BQ score between improved and not-improved patients. The cut-off with the best balance between sensitivity and specificity (Youden Index) was a change score of 14.5 at treatment 4/5 or treatment 10.

The third method of classifying patients as improved/not-improved was the distribution-based method. Using this method sixty eight percent of patients were improved at the treatment 4/5 and seventy three percent at the treatment 10. These values amounted to a reduction of at least 7.14 points at the treatment 4/5 and 7.44 points at the treatment 10. These values represented half the standard deviation of the mean change score for the BQ total score at treatment 4/5 and 10 respectively.

Table 22 demonstrates the level of agreement between the direct, anchor and distribution-based methods of classifying patients as improved. As with the back/leg pain cohort, the

strongest agreement was found between the anchor and distribution-based methods of categorising clinically significant improvement at treatment 4/5 and treatment 10. Moderate agreement was found between the PGIS and distribution as well as the PGIS and anchor-based method at treatment 10 with the weakest agreement between these methods at treatment 4/5.

Table 22. Agreement between three methods of assessing clinically significant improvement

	Methods Compared	Kappa Statistic	95% Confidence Interval
Treatment 4/5	*PGIS/Anchor-based	0.37	0.32-0.42
	*PGIS/Distribution	0.34	0.29-0.40
	Anchor/Distribution	0.59	0.55-0.64
Treatment 10	*PGIS/Anchor-based	0.48	0.34-0.61
	*PGIS/Distribution	0.47	0.32-0.61
	Anchor/Distribution	0.68	0.56-0.79

*= Patient Global Improvement Scale

3.2.3 STAGE 3: ASSOCIATIONS BETWEEN BASELINE VARIABLES

An analysis was carried out to look at the associations between potential predictors at baseline. Table 23 demonstrates the associations between the baseline variables using Chi-Squared analysis with the level of statistical significance set at p=0.05. Cells left blank on Table 23 indicate there was no significant association between baseline variables.

Table 23. Associations between potential predictor variables at baseline (*= chi-squared analysis p=<0.05)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Trauma (1)	-																		
Treatment Expectation (2)	*<0.01	-																	
Pain diagram (3)			-																
Gender (4)		*.05	*<0.01	-															
Employment status(5)	*.05			*<0.01	-														
Age (6)			*<0.01		*<0.01	-													
Inappropriate signs(7)		*.01	*<0.01	*<0.01			-												
Radiating pain(8)			*<0.01	*<0.01			*<0.01	-											
Widespread pain (9)		*.05	*<0.01	*.05	*<0.01		*<0.01	*<0.01	-										
Similar complaint (10)	*<0.01		*<0.01	*.05				*.02	*.04	-									
Pain Duration (11)	*.04	*<0.01			*<0.01	*.01	*<0.01		*<0.01		-								
Medication use (12)				*<0.01		<0.01*	*.01	*<0.01	*.02		*.03	-							
Work satisfaction (13)		*.05			*<0.01		*.03				*<0.01		-						
Work Expectation (14)		*<0.01			*<0.01	*.01	*<0.01		*<0.01		*<0.01		*<0.01	-					
Smoker (15)															-				
Alcohol (16)			*.01	*<0.01	*<0.01	*.01	*.03		*<0.01		*<0.01	*.03	*.01		*<0.01	-			
Physical activity (17)			*<0.01	*<0.01	*.01		*<0.01	*.02	*<0.01		*<0.01	*.03	*.01	*<0.01		*<0.01	-		
General health (18)	*.02	*.05	*<0.01	*<0.01	*<0.01		*<0.01		*<0.01	0.05	*<0.01	*.01	*<0.01	*<0.01		*<0.01	*<0.01	-	
Severity (BQ<30/70) (19)			*<0.01	*<0.01	*.03	*.02	*<0.01	*<0.01	*<0.01		*<0.01	*<0.01	*<0.01			*.01	*<0.01	*<0.01	-

Numbers in column one for potential predictors relate to numbers (potential predictors) in row one

3.2.4 STAGE 4: UNIVARIATE REGRESSION ANALYSIS OF POTENTIAL PREDICTORS

Univariate regression analysis was carried out for the individual baseline variables (potential predictors) and significant combined variables from Stage 3. These results can be found in Table B in appendix vii. This analysis was only performed for the indirect, anchor-based method of categorising improved patients as this method have been identified as the most suitable when looking for back pain predictors in earlier data analysis. Table 24 shows those single predictors that were significantly (p<0.1) associated with improvement at treatment 4/5 and treatment 10.

Table 24. Baseline variables significantly associated with improvement at treatment 4/5 and treatment 10

	Treatment 4/5			Treatment 10		
Variable	Odds Ratio	95% Confidence Interval	P value (significance<0.1)	Odds Ratio	95% Confidence Interval	P value (significance<0.1)
Chronicity <7 weeks	2.20	1.72 to 2.8	<0.01	3.07	1.50 to 6.29	<0.01
Medication use	1.59	1.21 to 2.08	<0.01			
Work Expectation	1.56	0.95 to 2.56	0.08			
Smoker	1.23	0.97 to 1.55	0.09	2.07	1.07 to 3.1	0.03
Employed	1.44	1.07 to 1.94	0.02			
Treatment Expectation	1.52	1.19 to 1.95	<0.01	2.75	1.41 to 5.37	<0.01
Age <50.1	1.26	0.96 to 1.66	0.10			
Total BQ<30	.15	0.11-to 0.19	<0.01	.36	0.18 to 0.7	<0.01

Table 25 on page 84 shows interaction baseline variables that were significant (p<0.05) from Stage 3 of the analysis that were similarly entered into a univariate regression analysis with improvement at treatment 4/5 and treatment 10 using the anchor-based method as the dependent variable. Those interactions that were significantly (p<0.05) associated with

improvement are shown. The significance level was set at p<0.05 because of the number of significant variables identified.

3.2.5 STAGE 5: MULTIVARIABLE LOGISTIC REGRESSION ANALYSIS FOR OUTCOME PREDICTORS

This stage of analysis involved entering the significant baseline variables from the univariate analysis in Stage 4 into a multivariable regression analysis in order to determine those variables that were independently associated with outcome when adjusting for all other variables in the model. This was done in three steps: first the single significant predictor variables, second the interactive significant variables and thirdly, those variables that remained significant (p<0.05) in the multivariable models from the first two steps were entered together into the final predictive multivariable model.

i. *PREDICTORS OF OUTCOME AT TREATMENT 4/5*

a. STEP 1-Single Predictors

The 8 single predictors from Stage 4 (Table 24) at a significance level of p<0.1 were entered into a multiple logistic regression model for improvement at treatment 4/5. Two single predictors remained significant in the multiple regression model (Table 26). Without the predictor model, fifty five percent of the patients were correctly classified as improved. Using this predictor model, seventy one percent of the patients were correctly classified as improved.

Table 25. Baseline interaction variables significantly associated with improvement at treatment 4/5 and treatment 10

Interaction Variables	Treatment 4/5			Treatment 10		
	Odds Ratio	95% Confidence Interval	P value (significance<0.05)	Odds Ratio	95% Confidence Interval	P value (significance<0.05)
TreatmentExpectationXTrauma	1.42	1.10 to 1.84	0.01	3.12	1.57 to 6.2	<0.01
EmployedXTrauma	1.46	1.02 to 2.08	0.04			
Chronicity<7weeksXTrauma	2.04	1.59 to 2.61	<0.01	2.12	1.09 to 4.12	0.03
TraumaX MedicationUse	1.40	1.1 to 1.79	0.01			
Baseline BQ<30/70XTrauma	0.28	0.22 to 0.36	<0.01			
TreatmentExpectationXGender	1.38	1.04 to 1.85	0.03	2.16	1.04 to 4.5	0.04
ExpectationXInapproriateSigns	1.40	1.08 to 1.82	0.01	2.30	1.15 to 4.58	0.02
Chronicity<7weeksXTreatmentExpectation	1.77	1.35 to 2.32	<0.01	3.29	1.60 to 6.78	<0.01
TreatmentExpectationXWork Normally	1.78	0.96 to 3.28	0.07	4.84	0.91 to 25.9	0.07
TreatmentExpectationXGeneralHealth	1.45	0.94-to 2.24	0.09			
Pain DiagramXGeneralHealth	1.64	1.09-to 2.45	0.02			
Baseline BQ<30/70XPainDiagram	0.43	0.33-to 0.56	<0.01			
GenderXMedicationUse	1.28	0.10 to 1.63	0.05			
Baseline BQ<30/70XGender	0.33	0.26 to 0.43	<0.01	0.42	0.21 to 0.86	0.02
EmployedXChronicity<7weeks	1.66	1.16 to 2.37	0.01			
EmployedXGeneralHealth	1.66	0.94 to 2.91	0.08			
Baseline BQ<30/70XEmployed	0.67	0.46 to 0.99	0.04			
ChronicityXAge	1.58	1.12 to 2.21	0.01			
AgeXMedicationUse	1.48	1.07 to 2.06	0.02			
Baseline BQ<30/70XAge	0.47	0.32 to 0.67	<0.01			
Chronicity<7weeksXInappropriate signs	1.98	1.55 to 2.53	<0.01	2.79	1.42 to 5.46	<0.01
InappropriateSignsXMedic.	1.31	1.03 to 1.68	0.03			
InappropriateSignsXWork Expectation				12.67	1.48 to 108.36	0.02
Baseline BQ<30/70XInappropriateSigns	0.22	0.17 to 0.29	<0.01	0.39	0.20 to 0.79	0.01
MedicationUSeXRadiatingPain	1.33	1.04 to 1.71	0.02			
Baseline BQ30/70XRadiatingPain	0.31	0.23 to 0.41	<0.01			
Chronicity<7weeksXWidespreadPain	2.22	1.74 to 2.84	<0.01	4<0.01	1.97 to 8.13	<0.01
WidespreadPainXMedicationUse	1.51	1.16 to 1.96	<0.01			
WidespreadPainXWorkExpectation	1.90	1.08 to 3.34	0.03	4.82	1.19 to 19.53	0.03
Baseline BQ<30/70XWidespreadPain	0.02	0.15 to 0.25	<0.01	0.44	0.23 to 0.86	0.02
Chronicity<7weeksXMedicationUse	1.62	1.41 to 1.85	<0.01	1.76	1.19 to 2.6	<0.01
Chronicity<7weeksXGeneralHealth	1.67	1.14 to 2.44	0.01			
Chronicity<7weeksXBaselineBQ<30/70	0.62	0.48 to 0.81	<0.01			
GeneralHealthXMedicationUSe	1.59	1.07 to 2.35	0.02			
Baseline BQ<30/70XJobSatisfaction	0.60	0.39 to 0.94	0.03			
Baseline BQ<30/70XAlcoholUse	0.36	0.26 to 0.49	<0.01			
BaselineBQ<30/70XPhysicalActivity	0.54	0.39 to 0.77	<0.01			

Table 26. Baseline variables independently associated with improvement at treatment 4/5

Predictor	Odds Ratio	95% Confidence Interval	P value(significance <0.05)
Chronicity (<7 weeks)	1.97	1.46 to 2.65	<0.01
Baseline BQ score(<30/70)	0.17	0.12 to 0.23	<0.01

b. STEP 2-Interactive Predictors

The 36 interactive predictors from Stage 4 (Table 25) at a significance level of p<0.05 were entered into a multivariable logistic regression model. Four interactive predictors remained significant (Table 27). The percentage of patients improved was fifty five percent prior to the modelling and seventy one percent were correctly classified as improved using this model.

Table 27. Baseline interactive variables independently associated with improvement at treatment 4/5

Predictor	Odds Ratio	95% Confidence Interval	P value (significance<0.05)
Baseline BQ(<30/70)XTrauma	0.61	0.38 to 0.99	0.04
Baseline BQ(<30/70)XInappropriate Signs	0.31	0.2 to 0.48	<0.01
PainDiagramXGeneral Health	2.74	1.41 to 5.32	<0.01
Chronicity(<7weeks)XInappropriate Signs	1.74	1.06 to 2.86	0.03

c. STEP 3-Final Model

The final model in stage 5 included all significant single (Table 26) and interactive (Table 27) predictors. The two single predictors and one interactive predictor remained significant (Table 28). This model increased the predicted percentage of patients who were correctly classified as improved at treatment 4/5 from fifty four percent to seventy three percent.

Table 28. Final model showing all variables independently associated with improvement at treatment 4/5

Predictor	Odds Ratio	95% Confidence Interval	P value (significance<0.05)
Chronicity<7weeks	1.91	1.10 to 3.31	0.02
Baseline BQ Score<30/70	0.17	0.09 to 0.03	<0.01
Pain DiagramXGeneral HealthPerception	2.07	1.32 to 3.27	<0.01

ii. PREDICTORS AT TREATMENT 10

a. STEP 1-Single Predictors

The 4 single potential predictors from Stage 4 (Table 24) at a significance level of p<0.1 from Stage 4 were entered into a multiple logistic regression model for improvement at treatment 10. All 4 predictors remained significant in the multiple regression model although the confidence intervals were wider in the results for treatment 10 as the number of subjects in this model was 142 compared to 938 patients at treatment 4/5 (Table 29). Without the predictor model, sixty four percent of the patients were correctly classified as improved. Using this predictor model, the number of patients correctly classified as improved to sixty seven percent.

Table 29. Baseline variables independently associated with improvement at treatment 10

Predictor	Odds Ratio	95% Confidence Interval	P value (significance<0.05)
Chronicity (<7 weeks)	3.66	1.65 to 8.11	<0.01
Smoker	2.36	1.10 to 5.04	0.03
TreatmentExpectation	2.45	1.19 to 5.03	0.01
Baseline BQ score(<30/70)	0.44	0.21 to 0.92	0.03

b. STEP 2-Interactive Predictors

The 14 interactive predictors from stage 4 (Table 25) at a significance level of p<0.05 were entered into a multivariable logistic regression model. Two interactive predictors remained significant (Table 30). The percentage of patients correctly classified as improved was sixty-five percent prior to the modelling and seventy-three percent using this model.

Table 30. Baseline interactive variables independently associated with improvement at treatment 10

Interactive Predictors	Odds Ratio	95% Confidence Interval	P value (significance<0.05)
TreatmentExpectationXTrauma	7.65	1.48 to 39.51	0.02
Chronicity<7weeksXInappropriate Signs	6.37	1.59 to 25.56	0.01

c. STEP 3-Final Model

The final model in stage 5 included all significant single (Table 29) and interactive (Table 30) baseline predictors. Two of the single predictors remained in the final model but the interactive predictors dropped out of this model (Table 31). This model increased the predicted percentage of patients who were correctly classified as improved at treatment 10 from sixty four percent to sixty eight percent.

Table 31. Final model showing all variables independently associated with improvement at treatment 10

Predictor	Odds Ratio	95% Confidence Interval	P value (significance<0.05)
Smoker	2.67	1.22 to 5.83	0.01
Baseline BQ Score (<30/70)	0.42	0.20 to 0.9	0.02

In conclusion, at treatment 4/5 and at treatment 10, the single predictor of a baseline total BQ score of less than 30/70 was a significant predictor of non-improvement in the final multivariable model (A total BQ score in excess of 30/70 was required to predict improvement). The single predictor of chronicity (less than 7 week history for the current episode) was significant in the final model for treatment 4/5 and of either having been or being a smoker was a significant single predictor in the final model at treatment 10. At treatment 4/5 patients had one significant interactive predictor in the final model (more than 8 shaded squares on the pain diagram interacting with the patients' general health perception being good or excellent) but there were no interactive predictors in the final model at treatment 10.

3.2.6 STAGE 6: PROPERTIES OF FINAL PREDICTION MODEL

i.   Treatment 4/5

Only those predictors significantly associated with improvement at treatment 4/5 (Table 28) were entered into the final model. The final stage of analysis calculated the final predictive properties of the model. The ability of the model to predict who would improve (sensitivity) was seventy eight percent, who would not improve (specificity) was sixty six and a half percent, the positive predictive value of the model (the chance an improved patient has been correctly classified) was seventy three percent and the negative predictive value (the chance that a patient who has not improved has been correctly classified) was seventy two percent. The area under the ROC curve gave the accuracy of the model for distinguishing between patients improved and not improved as categorised by the anchor-based method (Table 32) and was 0.76 demonstrating reasonably good ability to distinguish patients.

Table 32. Final predictive model of improvement at treatment 4/5 in neck/arm pain (with or without headache) patients

	Variables Tested	Odds Ratio and 95% Confidence Interval	1. Sensitivity (95% CI) 2. Specificity (95% CI) 3. Numbers correctly classified improved (95% CI) 4. Number correctly classified not improved (95% CI) 5. Area under the ROC (95% CI)
	Chronicity (<7weeks)	2.05(1.58 to 2.69)	
	Pain Diagram X General Health Perception	2.14 (1.37 to 3.35)	
	Baseline BQ (<30/70)	0.14 (0.15 to 0.18)	
FINAL MODEL			1. **77.8% (74.3 to 81%)** 2. **66.5% (62.3 to 70.5%)** 3. **73.2% (69.6 to 76.5%)** 4. **71.8% (67.6 to 75.7%)** 5. **0.76 (0.73 to 0.79)**

ii.   Treatment 10

Table 33 demonstrates the final prediction model at treatment 10 which included the two single predictors shown to be statistically significant for categorising improvement at treatment 10 from Table 31. The sensitivity and specificity at treatment 10 was sixty seven percent. Seventy four percent of the patients were correctly classified as improved using this model and fifty nine percent of patients were correctly classified as not improved using the final model. The ability of the model (AUC) to distinguish between improved and not-improved patients was 0.71 demonstrating reasonably good ability of this model to perform this role.

Table 33. Final predictive model of improvement at treatment 10 for neck/arm (with or without) headache patients

	Variables Tested	Odds Ratio and 95% Confidence Interval	1. Sensitivity (95% Confidence Interval) 2. Specificity (95% Confidence Interval) 3. Numbers correctly classified improved (95% Confidence Interval) 4. Number correctly classified not improved (95% Confidence Interval) 5. Area under the ROC (95% Confidence Interval)
	Smoker	1.94 (1.02 to 3.72)	
	Baseline BQ (<30/70)	0.26 (0.14-0.5)	
FINAL MODEL			**1. 67% (57.4 to 75.3%)** **2. 67.1% (55.7 to 76.8%)** **3. 74.2% (64.5 to 82%)** **4. 59% (48.3 to 69%)** **5. 0.71 (0.63 to 0.78)**

Chapter 4: Discussion

This study set out to investigate if it is possible to predict outcome in patients presenting to a chiropractic practice. In common with most studies of back and neck pain, this study population was young or early middle-aged, predominantly in work and with no gender bias. Although a wide array of musculoskeletal conditions were reported as the primary complaint, most patients had back and/or neck pain and these patients were the subject of this study. Moreover, in agreement with the literature, most patients in this study had an acute episode of a chronic problem, experiencing a recurrence of a long-standing complaint (Hestbeak et al 2003; Dionne et al 2008; Leijon et al 2009; Hayden et al 2010; Koes et al 2010; Bergström et al 2011).

A multitude of subjective outcome measures are available to the clinician for use with their patients, but there are limitations (Lemieux et al 2007; Revicki et al 2008). If one uses a generic measure such as the SF-36, one can compare different disease populations (Khorsan et al 2007; Scoggins & Patrick 2009; Dawson et al 2010). Although these generic outcome measures tend to be short, years of research have shown poor correlation with function calling into question their use in determining clinically meaningful change (Samsa et al 1999). On the other hand, disease-specific instruments such as the RMDQ (Roland & Morris 1983), the ODI (Fairbank et al 1980) and NDI (Vernon & Moir 1991) are widely used in clinical trials although they do not capture all the aspects of the BPS model. Using these instruments, practitioners are often surprised that the results of clinical trials are at odds with their clinical experience (Fletcher et al 1992; Bolton 2003).

This dilemma led to the development and validation of the BQ for back and neck pain patients in clinical trials and routine clinical chiropractic practice (Bolton & Breen 1999;

Bolton & Humphreys 2002). This outcome measure captures all the commonly-measured aspects of the BPS model and has been shown to be as valid, reliable and responsive as the more widely-used disease-specific instruments (Bolton & Breen 1999; Bolton & Humphreys 2002). Moreover, it has gained recognition internationally having been translated and validated in number of European languages (Malmqvist 2001; King 2002; Hartvigsen et al 2005; Martel et al 2009; Schmitt et al 2009). It is for these reasons, and that it is validated for both back and neck pain patients in chiropractic practice, that the BQ was chosen as the outcome measure for the purposes of this study.

The outcome that was measured in this study, irrespective of whether or not it was as a result of specific or non-specific treatment effects, was 'improvement'. The literature describes three methods of assessing clinically significant improvement. The 'direct method' dichotomises subjects based on the rating the patient chooses on a PGIS (Houweling 2010). Patients have to report they are 'improved' or 'much improved' on the PGIS to be reliably classified as improved. This PGIS as a method of classifying improvement is seldom used in the literature although it is frequently used as the 'anchor' for classifying patients using change scores on an outcome measure in anchor-based methodology (Copay et al 2007). In contrast, the distribution-based method uses a cut-off based on the distribution of group change scores on the outcome measure and tends to yield a cut-off considerably more lax than that derived from the anchor-based method (Norman et al 2003; Eton et al 2004; Kamath & Dueck 2005; Lemieux et al 2007; Revicki 2008; Houweling 2010; Turner et al 2010). As this method is statistical and looks at the change scores in groups of patients rather than individual change scores, the literature suggests it should be used to verify the results of anchor-based methodology rather than being the preferred method of assessing clinically significant improvement in its own right (Jordan et al 2006; Lauridsen et al 2006; Copay et al 2007; Revicki et al 2008; Terwee et al 2010).

In this study the distribution method required a change score on the BQ of around half the change score required to demonstrate clinically significant improvement in back and neck pain patients using anchor-based methodology. Although this less robust cut-off would have classified many more patients as improved, it is possible that this lax cut-off would have led to identifying baseline variables as predictors of improvement which were not valid, based on a change score that was not necessarily meaningful to patients (Norman et al 2003; Eton et al 2004; Houweling 2010; Turner et al 2010). Moreover, the direct method, using the PGIS, would have been simple to use but it lacks validity as a method in its own right because of recall bias on the part of the patient when assessing clinical significance and again could have led to invalid findings (Wyrwich & Wolinsky 2000; Copay et al 2007). As the overwhelming method of choice for assessing clinical significant improvement in outcome studies is the 'anchor-based method', this method was used for categorising improvement in patients for the purpose of defining outcome and subsequent identification of predictor variables, using a rigorous cut-off for improvement which was meaningful to patients.

In the post-treatment questionnaire used in this study, patients were also asked to rate satisfaction and whether or not treatment expectations had been met. Satisfaction with care was very high amongst this study population although it did not correlate with the findings of improvement using the anchor-based method. Satisfaction is not necessarily the same as outcome. Patient satisfaction merely tells us about the patient experience regarding the quality of care received with patients having an inherent tendency to respond favourably, particularly when completing these scales in the practice setting (Rosenfeld 1998). It has been shown that patients' perception of improvement is a multi-factorial and more complex than either alleviation of symptoms or improvement in function alone (Beaton et al 2001; Hush et al 2009).

This study identified prognostic factors in patients consulting chiropractors for back/leg and neck/shoulder pain patients with or without headaches. A distinction was made between short-term outcomes at treatment 4/5 and the outcome at treatment 10 for a smaller group of patients with more long-standing symptoms who required a more extensive course of treatment. This resulted in four final predictive models, one for back pain patients in the short-term at treatment 4/5, over a more prolonged period at treatment 10, for neck pain patients in the short-term at treatment 4/5 and over a longer period at treatment 10.

The final model for predicting the outcome of back pain patients with or without leg pain in the short-term demonstrated good sensitivity for identifying patients who were likely to improve under chiropractic care. The specificity of the model for correctly identifying those patients who would not improve was acceptable but not as good as the sensitivity. This model for back pain patients in the short-term was reasonably good in its ability to distinguish between those patients who would improve/not improve.

Similarly, the final predictive model for patients with back pain at treatment 10 demonstrated very good sensitivity for predicting patients who would improve with chiropractic treatment. Once again the specificity of the final model for predicting patients who would not improve was more modest but acceptable. The ability of this model to predict patients who would or would not improve at treatment 10 was good. These models compare well with results of a study with similar methodology for predicting improvement in back pain patients, although the predictors were different and their final predictive model had better specificity than sensitivity (Malmqvist et al 2008).

Good sensitivity for predicting patients who would improve was demonstrated by the final model for neck pain patients with or without headaches at treatment 4/5. This model demonstrated reasonably good specificity for predicting patients who would not improve.

The ability of the model to predict who would/would not improve in the short-term with neck pain patients was reasonably good. The properties of this final model bore remarkable similarity to the properties of the final model for back pain patients.

The final model for the ability of predicting patients with neck/shoulder pain with or without headaches who would improve at treatment 10 demonstrated equal and acceptable sensitivity and specificity. This model was good for classifying patients who improved and less useful but acceptable for classifying patients who did not improve. The ability of the model to distinguish between patients who would/would not improve was slightly lower than that for patients earlier in their treatment but was still acceptable.

These final models for patients in the back or neck cohort at treatment 4/5 demonstrated remarkably similar properties apart from patients' treatment expectation not seeming to have a bearing on outcome for patients with neck pain. A significant proportion of patients in this study reported pain in more than one site, illustrating the difficulty of investigating musculoskeletal disorders on the basis of regional pain. In spite of this, most musculoskeletal research in primary care confines itself to single pain sites, for example low back pain, even though inevitably a proportion of participants will have musculoskeletal pain elsewhere (Leboeuf-Yde et al 2004; Hill et al 2007; Leboeuf-Yde et al 2009). Very few prognostic factor studies have looked at patients with general musculoskeletal pain although it has been stated that there is commonality in predictors of outcome for most musculoskeletal conditions (Mallen et al 2007).

The properties of the final prediction models appear consistent with the published literature with regard to pain intensity and psychosocial distress, duration of symptoms, treatment expectation and the patients' general health perception although the baseline predictors in these studies were derived from outcome measures other than the BQ (Denison et al 2004;

Leboeuf-Yde et al 2004; Michaelson et al 2004; Bot et al 2005; Ozegovic et al 2009; Bergström et al 2010; Lindell et al 2010). This would suggest that the information collected from the BQ is at least of equal value for predicting what will happen to patients undergoing chiropractic treatment, without these effects being attributable to the treatment. There is evidence in the literature that pain intensity, psychological distress, fear avoidance, depression and activities of daily living are predictors in their own right (Shannon et al 2001; Denison et al 2004; Michaelson et al 2004; Hansson et al 2006; Hill et al 2007; Côté et al 2008; Henschke et al 2008; Keeley et al 2008; Jensen et al 2009; Lakke et al 2009; Lindell et al 2010; Ssavedra-Hernández et al 2011) although for the purposes of this study, these individual domains were not analysed separately. The BQ looks at all the commonly-measured domains of the BPS model although it was the total BQ score that was used rather than looking at each domain individually. Further analysis of these data could determine that some domains of the BQ are more important than others as predictors of outcome in back and neck pain patients.

It is believed that this study is the first time the area shaded on a pain diagram has been identified as a predictor of outcome in back and neck pain. Previous authors have limited the use of the pain diagram to describing the distribution of pain and for assessing hypochondriacal tendencies although these authors had concluded that pain diagrams were unhelpful as a measure of psychopathology and should be used for assessing pain distribution alone (Langworthy et al 2002; Leboeuf-Yde et al 2007).

Some of the interactive properties of the final prediction models were interesting but not easy to explain. The final model for the back pain cohort at treatment 10 found inappropriate markings on the pain diagram was associated with improvement, albeit that

this predictor was associated with a smaller area shaded on the pain diagram, a higher BQ score and regular alcohol consumption. It is unclear as to why inappropriate markings on a pain diagram would be associated with improvement in back pain patients. The penalty point method of grading inappropriate signs on a pain diagram is well established in the literature and has been shown to have good inter-rater reliability (Udén et al 1988; Ohnmeiss 2000; Lacey et al 2005). However, there is consensus in the literature that a pain diagram is not an adequate tool for assessing psychological distress and should be used for assessing pain distribution alone with separate validated instruments used for assessing psychological complications (Jensen & Karoly 1992; Ohnmeiss 2000; Voorhies et al 2007). As inappropriate marking on the pain diagram has been found to be of questionable value by these authors, it is uncertain if 'inappropriate markings on a pain diagram' is a clinically meaningful predictor of improvement in its own right. As the predictor was associated with improvement, could it be that certain acute back pain patients make inappropriate markings on pain diagrams at presentation because they are distressed and want the practitioner to know how bad the pain is, but this does not have a detrimental effect on treatment outcome?

Similarly, the literature has failed to find a clear link between alcohol consumption and back pain although the patients who regularly consumed alcohol have been associated with an improved outcome (Leboeuf-Yde 2000; Foster et al 2009; Ndetan et al 2009; Holmberg & Thelin 2010). One study found patients with back or neck pain who consumed alcohol on a regular basis were less likely to take sick leave although these findings were limited to female public sector workers (Skillgate et al 2009). The results of this study might go some way to supporting the theory that regular alcohol consumption is associated with an improvement although this predictor was not found to be significant for neck pain patients.

It is difficult to understand why neck pain patients had a better chance of improving at treatment 10 if they had been or were currently a smoker. There is some suggestion that being a smoker might at best be a weak risk factor for developing low back pain but the literature is unclear as to how smoking in neck pain sufferers could be associated with an improvement (Leboeuf-Yde et al 1996; Leboeuf-Yde et al 1998; Leboeuf-Yde 1999; Zvolensky et al 2009). It should be noted that the unexpected predictors of inappropriate marking on the pain diagram and regular alcohol consumption in back pain patients as well as being a smoker in the neck pain cohort were only found at treatment 10 when the sample size in this study was considerably smaller.

Although age and gender were assessed as potential predictors and are often cited in the literature as adverse predictors if the patient is older and female, this did not appear to be the case for this study population (Leboeuf-Yde et al 2004; Michaelson et al 2004; Enthoven et al 2006; Hansson et al 2006; Henschke et al 2008; Schellingerhout et al 2008; Carroll et al 2009; Axén et al 2011). Moreover, levels of education and work status (blue-collar versus white collar) are often mentioned in the literature as predictors of outcome in back and neck pain (Hill et al 2007; Rubinstein et al 2008; Carroll et al 2009; Bergström et al 2011). However, the demographic questions preceding the pre-treatment BQ do not enquire into levels of education and therefore these variables were not investigated in this study.

The majority of patients in this study had low back pain. This is consistent with the literature which reports low back pain being the most common site of musculoskeletal pain followed by neck pain (Leboeuf-Yde et al 2009). However, unlike most back pain studies, more than half of this study population experienced lower limb pain. The demographic and clinical questions that patients completed prior to the seven questions of the pre-treatment BQ in this study did not allow the researcher to distinguish between "nerve root" and "referred"

leg pain associated with low back pain. Pain radiating into the leg or "sciatica" is cited as an adverse prognostic factor for low back pain (Dunn & Croft 2004). The incidence of leg pain associated with back pain in the literature is variable ranging from twenty two percent to around forty percent (Leboeuf-Yde et al 2009; Schulz et al 2011). However, more recent evidence suggests that when one adjusts for other factors such as severity and duration of pain, poorer outcomes for back pain sufferers with leg pain are not necessarily explained by the presence of leg pain itself (Hill et al 2011). This is a source of reassurance as leg pain was not identified as a predictor of outcome for the back pain population despite there being a higher incidence of lower limb pain in this study population compared to most low back pain studies (Leboeuf-Yde et al 2009; Schulz et al 2011). Similarly in the neck pain cohort there was a large group (more than two-fifths) who reported radiating upper limb pain although this was not identified as a predictor of outcome.

Whilst the psychometric properties of the seven domains that explore the BPS model in the BQ have been investigated for validity, reliability and responsiveness, the demographic and clinical questions that precede the seven domains of the BQ have received less scrutiny (Bolton & Breen 1999; Juniper et al 2009). Return to work is cited in the literature as an important treatment outcome (Dunn & Croft 2004). Although work absenteeism has been cited by some authors as a predictor of outcome, the pre-treatment BQ does not enquire if the patient is on sick leave (Hansson et al 2006; Hill et al 2007; Henschke et al 2008; Keeley et al 2008; Jensen et al 2009). The post-treatment BQ does have a question asking patients if they are on sick leave but it was not used in the analysis as a potential predictor as this study was based on baseline variables rather than variables determined at treatment 4/5 and 10.

Work absenteeism was low, as one would expect in a study population who were paying for their own treatment, and job satisfaction was high and as such it is unlikely that sick leave

would have been a predictor in this study population. This supports the findings of some authors who found sick leave had little or nothing to do with the outcome of an episode of back or neck pain (Denison et al 2004; Michaelson et al 2004). The post-treatment questionnaire used in this study regarding return to work needs some refinement as patients who do not work in paid employment tended to leave this question blank as it did not apply to them. It is likely that most patients in this study sought treatment because they were trying to remain at work in their paid employment as it is usually only public sector workers that receive full pay for periods when they are on sick leave. One needs a carefully-worded question enabling a high response rate to ensure that the conclusions drawn from the data analysis are accurate.

The incidence of widespread pain in this study population was very low in back and neck pain patients. It has been found in numerous studies that several sites of pain are associated with a less favourable prognosis and the incidence cited in the literature varies between five and twenty one percent (Holm, Carroll, Cassidy, Skillgate & Ahlbom 2007; Cöster et al 2008; Staud 2009; Natvig, Ihlebaek, Grotle, Brage & Bruunsgaard 2010). Patients in this study were asked: 'Is your pain all over?' An affirmative response to this question was very low in both back and neck pain patients. This questions the validity of the question itself as we know patients who shaded greater areas on the pain diagram were associated with a less favourable outcome. A different response may have been obtained if the question was worded: 'Is the area of pain of your presenting complaint becoming considerably more widespread?' instead.

More than two-thirds of the back pain patients and just under two-thirds of the neck pain patients had had previous episodes of pain. Previous research suggests that prior history of episodes of back or neck pain have an adverse effect on treatment outcome although it was

not a predictor in this study population (Denison et al 2004; Michaelson et al 2004; Hansson et al 2006; Hill et al 2007; Henschke et al 2008; Keeley et al 2008; Carroll et al 2009; Jensen et al 2009; Bergström et al 2011). Moreover, more than half of the back pain patients and just under half of the neck pain patients had had more than seven weeks of pain in their current episode. It was found that patients with a shorter history of back and neck pain (less than seven weeks) had a better chance of improving in this cohort of patients supporting previous research (Denison et al 2004; Leboeuf-Yde et al 2004; Enthoven et al 2006; Hansson et al 2006; Hill et al 2007; Keeley et al 2008; Jensen et al 2009; Bergström et al 2010; Lindell et al 2010; Foster et al 2010). Although the patients in this study answered this question literally, relating how long their *current* episode of pain had lasted, this question can mislead the practitioner as most patients presenting in this cohort were not experiencing their first episode of back or neck pain and had a much longer history of musculoskeletal problems as determined when the case history was taken.

The result of this study support the current perception that back and neck pain patients seem to experience recurrent episodes of pain of varying intensity which they manage rather than cure (Leijon et al 2009). This would appear to be supported by the fact that less than two-thirds of our study population expected to recover or improve with around a third expecting to remain the same, and a small percentage expecting to get worse. One of the reasons the patient expectations in this study could be so modest is that they misunderstood the question as it asks: "How do you expect your condition to change in the next few weeks?" It could be that had the question been prefixed with: "With treatment, how do you expect...." that the answers patient gave would have been different. Since the data analysis was conducted at the beginning of 2010, patients who indicated that they expected to "stay the same" or "get worse" have been questioned further. With the exception of two patients,

it was clear patients answered the question as if they were not having treatment, despite completing the questionnaire in a chiropractic clinic prior to their initial consultation.

Despite the fact that so many patients had modest treatment expectations, those who did expect to "recover or improve" were associated with improvement, supporting the findings of previous research that found that positive treatment expectation was associated with a more favourable outcome (Bot et al 2005; Ozegovic et al 2009; Lindell et al 2010). Moreover, nearly all back and neck pain patients indicated that their treatment expectation had been met on the post-treatment BQ despite the findings that many patients had not experienced a clinically significant improvement. Once again this raises the issue of confusing patient satisfaction with that of treatment outcome.

In this study it was found that when the patient's general health perception was 'good or excellent' it was associated with an improved outcome for back pain patients. These findings support previous research that general health perceptions had an effect on treatment outcome although the importance of this predictor is probably overlooked by practitioners in clinical practice (Leboeuf-Yde et al 2009). If this finding is validated in future research, it will be advisable for clinicians to use this question either verbally during taking the case history or by getting patients to complete a questionnaire including a question about general health perceptions.

This study was a single cohort design and did not assume that the outcome for back and neck pain patients (improvement) was as a consequence of treatment. This study describes what happens to patients undergoing chiropractic treatment for back and neck pain and the prognostic factors associated with improvement for those patients. Studies without a comparison group allow no inferences to be drawn about associations and as such the predictors identified in this study should be subjected to rigorous analytical studies (Carey &

Boden 2003). It has been argued that whilst observational research does not replace the need for randomised clinical trials (RCT), a carefully conducted study retains many of the benefits of an RCT allowing health-care providers, payers and patients to evaluate current best practice and improve clinical decision making as these studies answer questions in the 'real world' (Horn, DeJong, Ryser, Vaezie & Teraoka 2005). Moons et al (2009) have stated the best design to answer prognostic questions is a cohort study although one can combine the intervention and control group where the intervention has been ineffective in an RCT or use the treatment variable as a separate predictor if the intervention was effective.

The design of this study was pragmatic in nature and no attempt was made to keep patients in the study for the purposes of follow-up. There were a sizeable number of subjects for whom there was no follow-up and as such those patients were not included in the analysis. It is possible that considerable bias could have been introduced into the results of this study by not having follow-up questionnaires for the patients who did not undergo at least four treatments although authors have found that patients lost to follow-up do not always introduce bias to the study (Solberg, Sørlie, Sjaavik, Nygaard & Ingebrigtsen 2011). Moreover, although there were a large number of subjects in this study, the final predictive models for patients who underwent ten treatments was based on a smaller cohort which may have accounted for some unexpected findings in the predictor models.

The multivariable models used to determine predictors of outcome help to reduce the chances of confounding (McKee et al 1999; Moons, Altman, Vergouwe & Royston 2009). External validity is the ability to generalise the results of the study and this can be done in observational research providing the new study population has a similar range of predictor values (McKee et al 1999; Moons et al 2009a). If the study populations vary, validation studies become necessary (Moon et al 2009). However, the confidence intervals were

reasonably narrow in this study, especially amongst the larger cohort at treatment 4/5. Although as many baseline variables as possible were considered as potential predictors, it is possible that there could have been oversights. As there are no questions about sick leave on the pre-treatment BQ, this was not considered as a potential predictor. Furthermore, although it was apparent that the back and neck pain patients were not mutually exclusive and that some of the subjects of this study had both back and neck pain, the presence of pain in both areas, which has been cited by some authors as a predictor, was not assessed as a potential predictor (Leboeuf-Yde et al 2004; Hill et al 2007; Schellingerhout et al 2008; Leboeuf-Yde et al 2009). This study was restricted to early outcomes while patients were still consulting. We do not know if the predictive models presented hold true in the long-term. As patients with back and neck pain experience exacerbations and remissions, and tend to seek treatment when the problem is at its worse, validation studies should ideally have longer follow-up periods.

All variables tested in this study were self-reported by patients in the chiropractic practice, which while relevant, may be subject to reporting bias. Respondent bias is therefore a potential limitation of this study. Many patients who attend chiropractic clinics are referred by existing patients or other health-care practitioners. If the referring patient had had a positive treatment experience, this would have given the new patient an element of 'expectation bias' (Licciardone & Russo 2006). Moreover, factors such as persuasion, patient-physician rapport and the patients' frequent unconscious desire to be acceptable or please the treating practitioner all introduce a level of bias (Licciardone & Russo 2006). Attempts were made to reduce this bias where patients did not complete the questionnaires in front of practitioners or staff and were in a separate room. As patients in this study were self-funders, not many patients would want to accept that they had wasted their money and derived no benefit from their treatment as it is known from the literature that self-funders

respond better than those patients whose treatment is paid for by a third party payer (Leboeuf-Yde et al 2009). Patient clustering bias, where certain practitioners attract particular conditions, can occur although this bias can be addressed through multiple logistic regression modelling (Resnik, Liu, Hart & Mor 2008). Recall bias is where the patient's current health status biases their recall of their previous state of health with regard to pain intensity and disability, which can lead to patients overestimating the improvement they have experienced (Menezes et al 2007).

This study has identified predictors of improvement for back and neck pain patients in one chiropractic clinic. Future studies will need to determine if the predictors derived in this study can be validated in other populations of back and neck pain patients before one can ascertain that the prediction models identified in this study are useful in clinical practice. However, an accurate prediction model is of no use if it does not change clinician behaviour (Moon et al 2009a). Follow-up over a more prolonged period of time may help clinicians to know whether these models are of use for predicting the outcome of treatment alone or whether they can help in predicting outcome over time. Once these models have been validated, impact studies quantifying the effect of using these prognostic models on clinicians' behaviour, patient outcome and cost effectiveness can be done.

Chapter 5: Conclusion

This study set out to assess improvement in patients with back and neck pain undergoing chiropractic treatment and in particular, to determine whether or not it is possible to use baseline clinical and demographic characteristics to predict the outcome of self-reported improvement. The clinical and demographic characteristics of the patients in this study were similar to other back and neck research populations. There was some variability in the numbers of patients classified as reliably improved depending on whether the direct, anchor-based or distribution-based method was used to calculate improvement as the outcome of interest. Based on statistical analysis, the anchor-based method was chosen as the method of assessing clinically significant improvement for the purposes of this study.

For back pain patients two final models for predicting improvement were calculated, one for improvement at treatment 4/5 and the second for improvement at treatment 10. These models had good predictive properties identifying patient who were improved but were less specific for identifying patients who did not improve. Having an acute episode and high levels of pain severity were independently associated with improvement. Reporting a smaller area on the pain diagram was also associated with improvement in back pain patients. In the short term at treatment 4/5, patients who had a good perception of their general health and a positive treatment expectation at baseline were associated with improvement. Patients who regularly consumed alcohol and made inappropriate markings on the pain diagram were associated with improvement at treatment 10 although these predictors with interactive with predictors one could more readily understand and interpret.

The final model for neck pain patients at treatment 4/5 demonstrated good predictive power for improvement and has greater specificity than the final model for back pain at

predicting which patients would not improve. The final model for neck pain at treatment 10 was slightly less sensitive for identifying improvement but demonstrated better specificity for identifying patients who would not improve than the final models for back and neck pain at treatment visit 4/5 and back pain at treatment visit 10. In common with back pain patients, a higher BQ score was an independent predictor of improvement for neck pain patients at treatment 4/5 and 10. Similarly, having an acute episode of pain, perception of good general health and reporting a smaller area on the pain diagram was associated with improvement in neck pain patients at treatment 4/5. Unexpectedly, neck patients who were smokers or had been smokers were associated with improvement at treatment 10 although this was not a predictor for improvement in neck pain patients earlier in their treatment or for back pain patients at any stage of their treatment.

It would appear that there is a level of commonality in predictors of outcome for patients undergoing chiropractic treatment with both back and neck pain. This confirms recent suggestions in the literature that predictors for musculoskeletal conditions are similar. It is believed that this study is the first to associate the area shaded on the pain diagram as a predictor of improvement in musculoskeletal conditions. If the final predictive models of this study can be validated in other back and neck pain cohorts, an impact study could be conducted to inform clinicians, patients and commissioners about the likely outcome and cost-effectiveness of patients undergoing chiropractic treatment for back and neck pain.

Chapter 6: Reflections on the Professional Doctorate

6.1 Preface

In my first year of undergraduate education in 1982 one could not help but be impressed by the patient handling and communication skills held by some chiropractors on the academic staff at the Anglo-European College of Chiropractic (AECC) but to a first year student, it was apparent that some of the staff held to belief systems about the effectiveness of chiropractic treatment that were not based on scientific reasoning or any current published literature. It is worth noting at this stage the evidence-base for chiropractic treatment was practically non-existent and it was not until 1986 that the AECC set up the first research department in the world, closely followed by some chiropractic colleges in the USA. I made an undertaking to myself in my first year that whilst I aspired to have similar patient handling and communication skills, I was determined to try and keep my academic knowledge current and relevant to clinical practice throughout my professional career.

In August of 1996 the AECC held a weekend course as a taster to the first post-graduate Masters' Degree programme for practising chiropractors. Despite having attended more than a hundred continual professional development (CPD) events since graduating in 1987, it was apparent that the way in which I was practising was becoming outdated, often based on dogma and at odds with the evidence starting to be published in the peer-reviewed literature for the management of back and neck pain. This was followed a month later by the first three-day weekend of the taught component of the degree. Professor Bolton introduced the concepts of reflective practice and set the first assignment where we had to reflect on what had taken us from graduates to the point of enrolment on the MSc programme and what we hoped to achieve though this self-directed learning process. There

was a strong emphasis on the rehabilitative exercises programmes emerging from The Czech Republic and Scandinavia with lectures and workshops on how these concepts had been implemented into clinical practice in the USA, still the home of the majority of chiropractors. There was a substantial module on research methodology, mostly taught by Professor Bolton.

On the first day Professor Bolton explained that when we graduated with our MSc which would take between three and five years, it was envisaged that there would be a professional doctorate programme available for those MSc graduates who would like to continue practice-based research to a doctoral level. I was one of two chiropractors who enrolled in the second cohort from our professional in the professional doctorate programme offered by the University of Portsmouth in 2002. In retrospect, I am not sure this was the right decision at that stage, just having finished five years of studying on the MSc programme.

The College of Chiropractors was established in 1999 and members of the profession, with greatly diverse undergraduate education, were invited to apply for membership. Relative disregard for undergraduate education was taken into account when the Court of Electors decided, during the 'grandfathering stage', whether the applicants would be accepted as members or fellows of the College. Members of the profession had to list all the post-graduate education courses they had attended in the previous decade, the professional journals to which they subscribed, leadership roles in the profession which they had undertaken and whether they held or were undertaking any post-graduate qualifications. There have been substantial changes in the way that professions are educated and organised, many of which require mandatory CPD and in some cases revalidation is required

although this process has recently been found to be disproportionate for the chiropractic profession (Scott, Brown, Lunt & Thorne 2004).

One of the requirements of chiropractors accepted for membership/fellowship of the College was that they undertake a minimum of 20/30 hours of CPD respectively. This requirement of CPD to retain College membership preceded the mandatory requirement for CPD made by the General Chiropractic Council, the regulator of the profession in more recent years. As I was one of the first 30 or so chiropractors who had been exposed to the idea of purposeful, self-directed learning through reflective practice through the MSc programme, I was appointed as the CPD tutor for the South West Faculty of the College of Chiropractors and found myself having to teach these concepts to the 140 College members in my faculty as well as overseeing the CPD records submitted to me on an annual basis. I have found it interesting to note that the handful of members in my faculty who found it difficult to grasp the concepts of reflective practice, where clinicians answer questions that arise in clinical practice (Zeiger 2004) and who struggled to submit their annual CPD portfolio were often the same names that have appeared before the Professional Conduct Committee of the General Chiropractic Council in the years that followed. In retrospect this has convinced me of the value of purposeful, self-directed CPD based on a deeper reflective process by the practitioner trying to identify learning needs based on weaknesses or gaps in one's knowledge-base.

6.2 Reflections on two components of the taught component of the Professional Doctorate

The first taught component was to make the doctoral students into good reflective practitioners. As this was the first taught component it was not immediately apparent how it fitted into the whole scheme. In my cohort there were 16 students, all in the healthcare

sector and only two of us were working outside the NHS. Without realising it, I think

chiropractors working in the independent sector, having to be professionally and financially

successful are all reflective practitioners to some extent. Although one is advised, there is no

requirement in law for a novice chiropractor to work for someone before setting up in

practice. Even those practitioners who do work for someone are often in a remote location

with little or no face-to-face contact with their principal. For this reason, a practitioner,

wishing to be viable in practice, will learn quickly from their mistakes and improve their

methods of practice through the reflective process.

The College of Chiropractors encourages a voluntary and anonymous reporting process to

identify patterns in practice and help the professional as a whole avoid making the same

mistakes that have been made in the past. I was one of the first chiropractors to make use of

this facility where a pregnant patient presented at my practice with neck and arm pain with

headaches. Imaging was not possible because of her pregnancy but later in the pregnancy an

MRI was performed which revealed the patient had Arnold Chiari II malformation, not

usually an indication for spinal manipulation. I wrote up this case study for the Publication

and Dissemination module of the PD and can report that although the patient had

neurosurgery and still experiences nuisance symptoms, she did not suffer any detrimental

side-effects from spinal manipulation.

In recent times members of the British Chiropractic Association have been subjected to a

mass complaint concerning the contents of their websites from a sceptic blogger. In the

preparation of my defence, a sound knowledge gained during the research methodology

module of the PD informed me of the limitations of case reports, case series and other non-

experimental methodologies, enabling me to mount a robust defence of the claims I had

made on my website. I was able to inform the Professional Conduct Committee why the RCT, although it removes the possibility of chance in outcome, does have limitations in that it does not account for the experiences of individual patients and does not inform the reader whether statistically significant changes are clinically meaningful to patients. Moreover, because of the inclusion/exclusion criteria, results from RCTs are not always easily generalisable to clinical practice (external validity).

As chiropractors do not have sufficient evidence to underpin many aspects of clinical practice, unless the profession as a whole is given help in learning how to systematically record the outcomes they observe in clinical practice and disseminate this information through lecturing and/or publishing these findings, the scope of practice for the profession is likely to become increasingly narrow. Following the completion of my MSc, where different methods of determining clinically significant improvement in back and neck patients were compared, I realised that the collection of data about the patient experiences was not something that should be done for research studies alone. Making use of patient-reported outcome measures in my practice became routine as I questioned whether these outcomes measures could be used not only to look at treatment outcome but whether they had use in predicting these outcomes?

This study has looked at patients undergoing chiropractic treatment over the last nine years. The results have been presented in two parts at the European Chiropractors Conference in London in 2010 and in Zürich in 2011. Public speaking had always seemed daunting but the opportunity to do two presentations to our multidisciplinary cohort in the taught component of the professional doctorate helped to overcome many of these barriers. I hope to have the opportunity to address other groups of chiropractors to encourage them start

using PROMS routinely in clinical practice and as a consequence improve patient outcomes. This in turn should lead to an opportunity for the findings of this paper to be validated in a different study population.

6.3 Reflections on the research component of the professional doctorate

I am finding the knowledge I have gained on outcomes whilst patients undergo chiropractic treatment in this study invaluable in communicating the likely outcome to patients with strong psychosocial considerations, who often have sub-optimal treatment outcomes. Patients either accept the explanation and change their coping strategies, taking more responsibility for their health or on occasion find it difficult to accept that there is not a purely biological explanation for their problem and become defensive with a few choosing to go away and consider other treatment options such as cognitive behaviour therapy to assist with their coping mechanisms. Most patients seem genuinely interested in the independent predictors identified in this study and can feel quite relieved when they are helped to comprehend how their attitudes and beliefs can affect recovery, especially when they report having consulted several practitioners without success in the past.

One of the observations when we were putting raw data onto the spreadsheet was the number of patients for whom we did not have follow-up questionnaires. We have implemented a computerised diary system which highlights every fifth visit which enables the receptionist to routinely ask the patient to complete a post-treatment BQ, resulting in less patients being lost to follow-up. I realise what a challenge it is going to be to encourage chiropractors to use and interpret the BQ in routine clinical practice as I seem to be having difficulty encouraging my colleagues to incorporate the predictors identified in this study in their clinical impression/diagnosis. Routine assessment of predictors could be invaluable in

helping patients not to have unrealistic expectations of chiropractic treatment and might in turn reduce complaints to regulatory bodies or professional indemnity insurers when their unrealistic expectations have not been met.

6.4 The future

A paper based related to the findings of this research project has recently been published (Bolton & Hurst 2011) (Appendix xiii) and I feel motivated to improve data collection to investigate whether supervised exercise programmes can enhance positive treatment outcomes for chronic/recurrent back and neck pain patients. I hope to have the opportunity to address groups of chiropractors to disseminate the results of this project and have accepted an invitation from the College of Chiropractors to present the findings of the recent publication to their members on 1 February 2012. Other chiropractors are doing research on developing the most useful subjective outcome measures for chiropractors and others are looking at treatment outcome predictors. I hope to be instrumental in helping clinicians to produce more practiced-based research. Very few practitioners in full-time practice have done PhDs or professional doctorates and whilst the latter lends itself to clinic-based research, practitioners need to be equipped with the tools to encourage this professional development. The current leadership of the British Chiropractic Association is committed to research and once I have finished this thesis, I hope to be able to play a more active role in helping this to happen.

REFERENCES

Ailliet, L., Rubinstein, S.M. & de Vet, H.C. (2010). Characteristics of chiropractors and their patients in Belgium. *Journal of Manipulative and Physiological Therapeutics,* 33 (8), 618-625.

Andersson, G.B.J. (1999) Epidemiological features of chronic low back pain. *Lancet*, 354, 581-585.

Anderson, J.J. (2007). Mean changes versus dichotomous definitions of improvement. *Statistical Methods in Medical Research* 16, 7-12.

Assendelft, W.J., Pfiefle, C.E. & Bouter, L.M. (1995). Chiropractic in The Netherlands: a survey of Dutch chiropractors. *Journal of Manipulative and Physiological Therapeutics,* 18 (3), 129-134.

Axén, I., Bodin, L., Bergstrom, G., Halasz, L., Lange, F., Lovgren, P.W. et al. (2011). Clustering patients on the basis of their individual course of low back pain over a six month period. *BMC Musculoskeletal Disorders* 12: 99. doi: 10.1186/1471-2474-12-99

Axén, I., Rosenbaum, A., Röbech, R., Wren, T. & Leboeuf-Yde, C. (2002). Can patients reactions to the first chiropractic treatment predict early favourable treatment outcome in persistent low back pain? *Journal of Manipulative and Physiological Therapeutics* 25 (7), 450-454.

Beaton, D.E., Boers, M. & Wells, G.A. (2002). Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 14, 109-114.

Beattie, P.F., Nelson, R., Murphy, D.R. (2011). Development and Preliminary Validation of the MedRisk Instrument to Measure Patient Satisfaction with Chiropractic. *Journal of Manipulative and Physiological Therapeutics* 34 (10), 23-36.

Bendix, A.F., Bendix, T. & Haestrup, C. (1998). Can it be predicted which patients with chronic low back pain should be offered tertiary rehabilitation in a functional restoration program? A search for demographic, socioeconomic and physical predictors. *Spine* 23, 1775-1783.

Bergström, G., Bergström, C., Hagberg, J., Bodin, L. & Jensen, I. (2010) A 7-year follow-up of multidisciplinary rehabilitation among chronic neck and back pain patients. Is sick leave outcome dependent on psychologically derived patient groups? *European Journal of Pain* 14 (4), 426-433.

Bergström, C., Hagberg, J., Bodin, L., Jensen, I. & Bergström, G. (2011). Using a psychosocial subgroup assignment to predict sickness absence in a working population with neck and back pain. *Biomed Central Musculoskeletal Disorders*, 12: 81. http://www.biomedcentral.com/1471-2474/12/81

Bergström, C., Jensen, i., Hagberg, J., Busch, H. & Bergström, G. (2011). Effectiveness of different interventions using a psychosocial subgroup assignment in chronic neck and back pain patients: a 10-year follow-up. *Disability and Rehabilitation*, 1-9, Early Online. Downloaded from www.informahealthcare.com by 94.193.104.143.

Bolton, J.E. (1993). Methods of assessing low back pain and related psychological factors: a literature review. *European Journal of Chiropractic* 41, 31-38.

Bolton, J.E. (1994). Evaluation of treatment in back pain patients: clinical outcome measures. *European Journal of Chiropractic* 42, 29-40.

Bolton, J.E. (1997). On the responsiveness of outcome measures. *European Journal of Chiropractic* 45, 5-8.

Bolton, J.E. (2003). Research practice and clinical practice: as far apart as ever? *Clinical Chiropractic* 6, 92-94.

Bolton, J.E. (2004). Sensitivity and Specificity of Outcome Measures in Patients with Neck Pain: Detecting Clinically Significant Improvement. *Spine* 29 (21), 2410-2417.

Bolton, J. E. & Breen, A.C. (1999). The Bournemouth Questionnaire: A short-form comprehensive outcome measure. I. Psychometric properties in back pain patients. *Journal of Manipulative and Physiological Therapeutics* 22 (8), 503-510.

Bolton, J.E. & Humphreys, B.K. (2002). The Bournemouth Questionnaire: A short-form comprehensive outcome measure. I. Psychometric properties in neck pain patients. *Journal of Manipulative and Physiological Therapeutics* 25, 141-148.

Bolton, J.E. & Hurst, H.C. (2011). Prognostic factors for short-term improvement in acute and persistent musculoskeletal pain consulters in primary care. *Biomed Central Chiropractic & Manual Therapies* 19: 27. doi: 10.1186/2045-709X-19-27

Bolton, J.E. & Wilkinson, R.C. (1998). Responsiveness of Pain Scales: A Comparison of Three Pain Intensity Measures in Chiropractic Patients. *Journal of Manipulative and Physiological Therapeutics* 21 (1), 1-7.

Borge, J.A., Leboeuf-Yde, C. & Lothe, J. (2001). Prognostic Values of Examination Findings in Patients with Chronic Low Back Pain Treated Conservatively: A Systematic Review. *Journal of Manipulative and Physiological Therapeutics* 24 (4), 292-295.

Borkan, J.M., Koes, B., Reis, S. & Cherkin, D.C. (1998). A report from the Second International Forum for Primary Care Research on Low Back Pain. Re-examining Priorities. *Spine* 23, 1992-1996.

Bot, S.D.M., van der Waal, J.M., Terwee, C.B., van der Windt, D.A.W.M., Scholten, R.J.P.M., Bouter, L.M. et al. (2005). Predictors of Outcome in Neck and Shoulder Symptoms. *Spine* 30 (16): E459-E470.

Bousema, E.J., Verbunt, J.A., Seelen, H.A.M., Vlaeyen, J.W.S. & Knottnerus, J. A. (2007). Disuse and physical deconditioning in the first year after the onset of back pain. *Pain* 130, 279-286.

Breen A. (2010) Annual Evidence Update on CAM for Low Back Pain-Chiropractic. NHS Evidence-Complementary and alternative medicine-2010. http://www.library.nhs.uk/cam/viewResource.aspx?resid=397040&code=9681b&d345 retrieved 15/12/2010.

Bronfort, G., Haas, M., Evans, R.L. & Bouter, L.M. (2004) Efficacy of spinal manipulation and mobilization for low back pain and neck pain: a systematic review and best evidence synthesis. *The Spine Journal* 4, 335-356.

Bronfort, G., Haas, M., Evans, R., Leininger, B. & Triano J. (2010). Effectiveness of Manual Therapies: the UK evidence report. *Biomed Central Chiropractic and Osteopathy* 18:3. doi: 10.1186/1746-1340-18-3

Buitenhuis, J., de Jong, P.J., Jaspers, J.P. & Groothoff, J.W. (2008) Catastrophizing and causal beliefs in whiplash. *Spine* 33 (22), 2427-2433.

Carey, T.S. & Boden, S.D. (2003). A Critical Guide to Case Series Reports. *Spine* 28 (3), 1631-1634.

Carroll, L.J., Cassidy, J.D. & Côté, P. (2004). Depression as a risk factor for onset of an episode of troublesome neck and low back pain. *Pain* 107, 134-139.

Carroll, L.J., Hogg-Johnson, S., Côté, P., van der Velde, G., Holm, L.M., Carragee, E.J. et al. (2009) Course and Prognostic Factors for Neck Pain in Workers. Results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and its Associated Disorders. *Journal of Manipulative and Physiological Therapeutics* 32 (2S), S108-S116.

Chen, Y., Li, L., Chen, X., Zhang, Q., Wong, W., Li, Y. & Yang, D. (2008). Ovarian volume and follicle numbers in the diagnosis of polycystic ovarian syndrome in Chinese women. *Ultrasound Obstetrics and Gynaecology* 32 (5), 700-703.

Chenot, J.F., Scherer, M., Becker, A., Donner-Banzhoff, N., Baum, E., Leonhardt, C. et al. (2008). Acceptance and perceived barriers of implementing a guideline for managing low back in general practice. *Implementing Science* 3:7 doi 10.1186/1748-5908-3-7

Cheshire, A., Polley, M., Peters, D & Ridge, D. (2011). Is it feasible and effective to provide osteopathy and acupuncture for patients with musculoskeletal problems in a GP setting? A service evaluation. *BioMed Central Family Practice*12: 49. http://www.biomedcentral.com/1471-2296/12/49

Christensen, L. & Mendoza, J.L. (1986). A method of assessing change in a single subject: an alteration of the RC index. *Behavioural Therapy* 17, 305-308.

Coons, S.J. (2008). Describing Response Scales in Patient-Reported Outcome Measures: Terms Matter. *Clinical Therapeutics* 30 (12), 2408-2409.

Copay, A.G., Glassman, D.D., Subach, B.R., Berven, S., Schuler, T.C. & Carreon, L.Y. (2008). Minimal clinically important differences in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study Questionnaire Short Form 36, and Pain Scales. *The Spine Journal* 8, 968-974.

Copay, A.G., Subach, B.R., Glassman, S.D., Polly, D.W. & Schuler, T.C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 7, 541-546.

Cöster, L., Kendall, S., Gerdle, B., Henriksson, C., Henriksson, K.G. & Bengtsson, A. (2008). Chronic widespread musculoskeletal pain-a comparison of those who meet criteria for fibromyalgia and those who do not. *European Journal of Pain* 12 (5), 600-610.

Côté, P., van der Velde, G., Cassidy, J.D., Carroll, L.J., Hogg-Johnson, S., Holm, L.W. et al. (2008). The burden and determinants of neck pain in workers: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 15 (33 Supplement 4), S60-74.

Crosby, R.D., Kolotkin, R.L. & Williams, G.R. (2003). Defining clinically meaningful change in health-related quality of life. Journal of Clinical Epidemiology 56, 395-407.

Dahl, T.H. (2002). International Classification of Functioning, Disability and Health: An introduction and discussion of its potential impact on rehabilitation services and research. *Journal of Rehabilitative Medicine* 34, 201-204.

Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C. & Carr, A.J. (2010). Routine use of patient reported outcome measures in healthcare settings. *British Medical Journal* 340, 464-467.

Denison, E., Åsenlöf, P. & Lindberg P. (2004). Self-efficacy, fear avoidance, and pain intensity as predictors of disability in subacute and chronic musculoskeletal pain patients in primary care. *Pain* 111, 245-252.

Department of Health. (2006). The Musculoskeletal Services Framework. A joint responsibility: doing it differently. http://www.nhs.uk/Asset.askx?path=/orthopaedics/Muskuloskeltal%20Framework_241007.pdf. Retrieved updated version 21/11/2009

Derebery, J., Giang, G.M., Gatchel, R.J., Erickson, K. & Fogarty, T.W. (2009). Efficacy of a patient-education booklet for neck-pain patients with workers' compensation: a randomized controlled trial. *Spine* 34 (2), 206-213.

De Vet, H.C.W., Terluin, B., Knol, D.L., Roorda, L.D., Mokkink, L.B., Ostelo, R.W.J.G. et al. (2010). Three ways to quantify uncertainty in individually applied "minimally important change" values. *Journal of Clinical Epidemiology* 63 (1), 37-45.

De Vet, H.C., Terwee, C.B., Ostelo, R.W., Beckerman, H., Knol, D.L. & Bouter, L.M. (2006). Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Quality Life Outcomes* 4: 54.

Deyo, R.A., Andersson, G., Bombardier, C., Cherkin, D.C., Keller, R.B., Lee, C.K. et al. (1994). Outcome Measures for studying patients with low back pain. *Spine* 19 (185), 2032S-2036S.

Dionne, C.E., Dunn, K.M., Croft, P.R., Nachemson, A.L., Buchbinder, R., Walker, B.F. et al. (2008) A Consensus Approach Towards the Standardisation of Back Pain Definitions for the Use in Prevalence Studies. *Spine* 33 (1), 95-103.

Dunn, K.M. & Croft, P.R. (2004). Epidemiology and natural history of low back pain. *Europa Medicophysica* 40 (1), 9-13.

Eltayeb, S., Staal, J.B., Hassan, A. & de Bie, R.A. (2007). Work Related Risk Factors for Neck, Shoulder and Arm Complaints: A Cohort Study Among Dutch Computer Office Workers. *Journal of Occupational Rehabilitation* 19, 315-322.

Eltayeb, S., Staal, J.B., Kennes, J., Lamberts, P.H.G. & de Bie, R.A. (2007). Prevalence of complaints of arm, neck and shoulder among computer office workers and psychometric evaluation of a risk factor questionnaire. *Biomed Central Musculoskeletal Disorders,* 8, 68. http://www.biomedcentral.com/1471-2474/68

Enthoven, P., Skargren, E., Carstensen, J. & Öberg, B. (2004). Clinical Course in Patients Seeking Primary Care for Back or Neck Pain: A Prospective 5-Year Follow-Up of Outcome and Health Care Consumption with Subgroup Analysis. *Spine* 29 (21), 2458-2465.

Enthoven, P., Skargren, E., Carstensen, J. & Öberg, B. (2006). Predictive factors for 1-year and 5-year outcome for disability in a working population of patients with low back pain treated in primary care. *Pain* 122, 137-144.

Eton, D.T., Cella, D., Yost, K.J., Yount, S., Peterman, A.H., Neuberg, D.S. et al. (2004). A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *Journal of Clinical Epidemiology* 57, 898-910.

Fairbank, J.C.T, Couper, J., Davies, J.B. & O'Brien, J.P. (1980). The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy* 66, 271-273.

Feleus, A., Bierma-Zeinstra, S.M.A., Miedma, H.S., Bernsen, R.M.D., Verhaar, J.A.N. & Koes, B.W. (2008). Incidence of non-traumatic complaints of arm, neck and shoulder in general practice. *Manual Therapy*, 13, 426-433.

Fletcher, A., Gore, S., Jones, D., Fitzpatrick, R., Spiegelhalter, D. & Cox, D. (1992). Quality of Life Measures in Health Care II: design, analysis and interpretation. *British Medical Journal* 305, 1145-1148.

Fluss, R., Faraggi, D. & Reiser B. (2005). Estimation of the Youden Index and its associated cut-off. *Biomedicine Journal* 47 (4), 458-472.

Flynn, T., Fritz, J., Whitman, J., Wainner, R., Magel, J., Rendeiro, D. et al. (2002). A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine* 27(24), 2835-2843.

Fritz, J., Childs, J.D. & Flynn, T.W. (2005). Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. *Biomed Central Family Practice* 6: 29 doi: 10.1186/1471-2296-6-29

Foster, N.E., Dziedzic, K.S., van der Windt, D., Fritz, J.M. & Hay, E.M. (2009). Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. *Biomed Central Musculoskeletal Disorders,* 10, 3. http://www.biomedcentral.com/10.1186/1471-2474/10/3

Foster, N.E., Hill, J.C. & Hay, E.M. (2011) Subgrouping patients with low back pain in primary care: Are we getting any better? *Manual Therapy* 16, 3-8.

Foster, N.E., Thomas, E., Bishop, A., Dunn, K.M. & Main, C.J. (2010). Distinctiveness of psychological obstacles to recovery in low back pain patients in primary care. *Pain* 148, 398-406.

Garner, M.J., Birmingham, M., Aker, P., Moher, D., Balon, J., Keenan, D. et al. (2008). Developing Integrative Primary Healthcare Delivery: Adding a Chiropractor to the Team. *Explore* 4 (1), 18-24.

Gatchel, R.J. & Mayer, T.G. (2010). Testing minimal clinically important difference: consensus or conundrum? *The Spine Journal* 10, 321-327.

Gay, R.E., Madson, T.J. & Cieslak, K.R. (2007). Comparison of the Neck Disability Index and the Neck Bournemouth Questionnaire in a Sample of Patients with Chronic Uncomplicated Neck Pain. *Journal of Manipulative and Physiological Therapeutics* 30 (4), 259-262.

Great Britain. Government Statistics Office. (1999). The Prevalence of Back Pain in Great Britain in 1998. Bulletin 1999/18 ISBN 1 841820563

Greene, B.R., Smith, M., Allareddy, V. & Haas M. (2006) Referral patterns and attitudes of Primary Care Physicians towards chiropractors. *Biomed Central Complementary and Alternative Medicine* 6: 5 doi: 10.1186/1472-6882-6-5

Guyatt, G.H. & Jaeschke, R.J. (1997). Reassessing quality of life instruments in the evaluation of new drugs. *Pharmacoeconomics* 12 (6), 621-626.

Haas, M., Spegman, A., Peterson, D., Aickin, M. & Vavrek D. (2010). Dose-Response and Efficacy of Spinal Manipulation for Chronic Cervicogenic Headache: A Pilot Randomised Controlled Trial. *The Spine Journal,* 10 (2), 117-128.

Hageman, W.J.J.M. & Arrindel, W.A. (1999). Establishing clinically significant change: increment of precision and the distinction between individual and group levels of analysis. *Behaviour Research and Therapy* 37, 1169-1193.

Hagen, E. M., Svensen, E. & Eriksen, H.R. (2005). Predictors and Modifiers of Treatment Effect Influencing Sick Leave in Subacute Low Back Pain Patients. *Spine* 30 (4), 2717-2723.

Hägg, O., Fritzell, P. & Nordwall A. (2003). The clinical importance of change scores in outcome scores after treatment for chronic low back pain. *European Spine Journal* 12 (1), 12-20.

Hall, H., McIntosh, G. & Boyle, C. (2009). Effectiveness of a low back pain classification system. *The Spine Journal* 9, 948-657.

Hamberg-van Reenen, H.H., Ariëns, G.A.M., Blatter, B.M., van Mechelen, W. & Bongers, P.M. (2007). A systematic review of the relationship between physical capacity and future low back and neck/shoulder pain. *Pain* 130, 93-107.

Hancock, M., Herbert, R.D. & Maher, C.G. (2009). A Guide to Interpretation of Studies Investigating Subgroups of Responders to Physical Therapy Interventions. *Physical Therapy* 89 (7), 698-704.

Hansson, E., Hansson, T. & Jonsson, R. (2006). Predictors for work ability and disability in men and women with low-back or neck problems. *European Spine Journal* 15 (6), 780-793.

Harkness, E.F., Macfarlane, G.J., Silman, A.J. & McBeth, J. (2005). Is musculoskeletal pain more common now than 40 years ago?: two population-based cross sectional studies. *Rheumatology*, 44, 890-895.

Hartvigsen, J., Lauridsen, H., Ekström, S., Nielsen, M.B., Lange, F., Kofoed, N. et al. (2005). Translation and validation of the Danish version of the Bournemouth Questionnaire. *Journal of Manipulative and Physiological Therapeutics* 28 (6), 402-407.

Hayden, J.A., Dunn, K.M., van der Windt, D.A. & Shaw, W.S. (2010). What is the prognosis of back pain? *Best Practice and Research Clinical Rheumatology* 24, 167-179.

Hays, R.D. & Woolley, J.M. (2000). The concept of clinically meaningful difference in health-related quality of life research. *Pharmacoeconomics* 18, 419-423.

Henschke, N., Maher, C., Refshauge, K.M., Herbert, R.D., Cumming, R.G., Bleasel, J. et al. (2008). Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *British Medical Journal* 337:a171 doi:10.1136/bmj.a171

Hestbaek, L., Lebouef-Yde, C. & Manniche, C. (2003). Low back pain: What is the long-term course? A review of studies of general populations. *European Spine Journal* 12, 149-165.

Hildebrandt, J., Ursin, H., Mannion, A.F., Airaksinen, O., Brox, J.I., Cedraschi, C. et al. (2004). European Guidelines for the Management of Chronic Non-Specific Low Back Pain. *European Commission Research Directorate General*.
http://www.backpaineurope.org/web/html/wg2_results.html

Hill, J.C., Lewis, M., Sim, J., Hay, E.M. & Dziedzic, K. (2007). Predictors of poor outcome in patients with neck pain treated by physical therapy. *Clinical Journal of Pain* 23 (8), 683-690.

Hill, J.C., Konstantinou, K., Egbewale, B.E., Dunn, K.M., Lewis M. & van der Windt D. (2011). Clinical Outcomes Among Low Back Pain Consulters With Referred Leg Pain In Primary Care. *Spine* Published Ahead of Print. doi: 10.1097/BRS.0b013e31820712bb

Hillman, M., Wright, A., Rajaratnam, G., Tennant, A., Chamberlain, M.A. (1996) Prevalence of low back pain in the community: implications for service provision in Bradford, UK. *Journal of Epidemiology and Community Health*, 50, 347-352.

Holm, L.W., Carroll, L.J., Cassidy, J.D., Skillgate, E. & Ahlbom, A. (2007). Widespread pain following whiplash-associated disorders: incidence, course and risk factors. *Journal of Rheumatology* 34 (1), 193-200.

Holmberg, S.A.C. & Thelin, A.G. (2010) Predictors of sick leave owing to neck or low back pain: A 12-year longitudinal cohort study in a rural male population. *Annals of Agricultural Environmental Medicine* 17, 251-257.

Holtermann, A., Hansen, J.V., Burr, H. & Søgaard, K. (2010) Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. *Scandinavian Journal of Environmental Health* 36 (1), 34-41.

Horn, S.D., De Jong, G., Ryser, D.K., Vaezie, P.J. & Teraoka, J. (2005). Another look at observational studies in rehabilitation research: Going beyond the holy grail of the randomised controlled trial. *Archive of Physical Medicine Rehabilitation* 86 (2), S8-S15.

Hosmer, D.W. & Lemeshow, S. (2000). *Applied Logistic Regression*. Second Edition. New York: Wiley.

Houweling, T.A.W. (2010). Reporting Improvement from patient-reported outcome measures: A review. *Clinical Chiropractic* 13, 15-22.

Hurst, H.C. (2001). *Agreement between anchor-based method and distribution-based estimates of clinically important differences*. Unpublished MSc Clinical Chiropractic Dissertation. University of Portsmouth, Portsmouth.

Hurst, H. & Bolton, J. (2004). Assessing the Clinical Significance of Change Scores Recorded on Subjective Outcome Measures. *Journal of Manipulative and Physiological Therapeutics* 27, 26-35.

Hush, J.M., Michaleff, Z., Maher, C.G., & Refshauge, K. (2009). Individual, physical and psychological risk factors for neck pain in Australian office workers: a 1-year longitudinal study. *European Spine Journal* 18, 1532-1540.

Jacobson, N.S., Follette, W.C. & Revenstorf, D. (1984). Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behavioural Therapy* 15, 336-352.

Jacobson, N.S. & Traux, P. (1991). Clinical Significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 59, 12-19.

Jaeschke, R., Singer, J. & Guyatt, G. (1989). Measurement of health status, ascertaining the minimal clinical important difference. *Controlled Clinical Trials* 10, 407-415.

Jensen, T.S., Albert, H.B., Sorensen, J.S., Manniche, C. & Leboeuf-Yde, C. (2007). Magnetic Resonance Imagining Findings as a predictor of clinical outcome in patients with sciatica receiving active conservative treatment. *Journal of Manipulative and Physiological Therapeutics* 30 (2), 98-108.

Jensen, I.B., Busch, H., Bodin, L., Hagberg, J., Nygren, Å. & Bergström, G. (2009). Cost effectiveness of two rehabilitation programmes for neck and back pain patients: A seven year follow-up. *Pain* 142, 202-208.

Jensen, M.P. & Karoly, P. (1992). Self-Report Scales and Procedures. In Turk, D.C. & Melzack, R. (Eds). *Handbook of Pain Assessment,* pp. 147-148. Guildford Press.

Jolles, B.M., Buchbinder, R. & Beaton, D.E. (2005). A study compared nine patient-specific indices for musculoskeletal disorders. *Journal of Clinical Epidemiology* 58, 791-801.

Jordan, K., Dunn, K.M., Lewis, M. & Croft, P. (2006). A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. *Journal of Clinical Epidemiology* 59, 45-52.

Jordan, K.P., Kadam, U.T., Hayward, R., Porcheret, M., Young, C., & Croft, P. (2010). Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study [Electronic version]. *Biomed Central Musculoskeletal Disorders*, 11:144. http://www.biomedcentral.com/1471-2474/11/144

Juniper, M., Le, T.K. & Mladsi, D. (2009). The epidemiology, economic burden and pharmacological treatment of chronic low back pain in France, Germany, Italy, Spain, and the UK: a literature-based review. *Expert Opinion on Pharmacotherapy*, 10 (16), 2581-2592.

Kamath, C. & Dueck, A. (2005). Incorporating Clinical Significance into a Study. *Current Problems in Cancer* 29 (6), 306-316.

Karjalainen, K., Malmivaara, A., Mutanen, P., Roine, R., Hurri, H. & Pohjolainen, T. (2004). Mini-Intervention for Subacute Low Back Pain. *Spine* 29 (10), 1069-1076.

Keeley, P., Creed, F., Tomenson, B., Todd, C., Borgin, G. & Dickens, C. (2008). Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain. *Pain* 135, 142-150.

Keijers, E., Feleus, A., Miedema, H.S., Koes, B.W. & Bierma-Zeinstra, S.M.A. (2010). Psychological factors predicted non-recovery in both specific and nonspecific diagnoses at arm, neck and shoulder. *Journal of Clinical Epidemiology,* 63, 1370-1379.

Khorsan, R., Coulter, I.D., Hawk, C. & Choate, C. G. (2008). Measures in Chiropractic Research: Choosing Patient-Based Outcome Assessments. *Journal of Physiological and Manipulative Therapeutics* 31 (5), 355-375.

King, R. (2002). *The Spanish (Castilian) version of the Bournemouth Questionnaire: reliability, validity and responsiveness*. Unpublished MSc Clinical Chiropractic Dissertation, University of Portsmouth, Portsmouth.

Koes, B.W., van Tulder, M., Chung-Wei, C.L., Macedo, L.G., McAuley, J. & Maher, C. (2010). An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *European Spine Journal* 19: 2075-2094.

Koes BW, van Tulder MW, Ostello R, Burton AK & Waddell G. (2001). Clinical Guidelines for the Management of Low Back Pain in Primary Care: An International Comparison. *Spine* 26 (22), 2504-2514.

Koleck, M., Mazaux, J-M., Rascle, N. & Bruchon-Schweitzer, M. (2006). Psycho-social factors and coping strategies as predictors of chronic evolution and quality of life in patients with low back pain: A prospective study. *European Journal of Pain* 10, 1-11.

Kongsted, A. & Lebouef-Yde C. (2010). The Nordic back pain subpopulation program: Can low back pain patterns be predicted from the first consultation with a chiropractor? A longitudinal pilot study. *Biomed Central Chiropractic and Osteopathy* 18:8. http://www.chiropracticandosteo.com/content/18/1/8

Krebs, E.E., Bair, M.J., Damush, T.M., Tu, W., Wu, J. & Kroenke, K. (2010). Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. *Medical Care* 48 (11), 1007-1014.

Lacey, R.J., Lewis, M., Jordan, K., Jinks, C. & Sim, J. (2005). Interrater reliability of scoring of pain drawings in self-report health survey. *Spine* 30 (16), E455-458.

Lakke, S.E., Soer, R., Takken, T. & Reneman, M.F. (2009). Risk and prognostic factors for non-specific musculoskeletal pain: a synthesis of the evidence from systematic reviews classified into ICF dimensions. *Pain* 147 (1-3), 153-164.

Langworthy, J. & Breen, A.C. (2007). Psychosocial factors and their predictive value in chiropractic patients with low back pain: a prospective inception cohort study. *Biomed Central Chiropractic and Osteopathy* 15, 5. doi:10.1186/1746-1340-15-5

Langworthy, J.M., Breen, A.C., Vogel, S. & Collier, R. (2002) Third Prize: Chiropractic and the national health care system: a basis for partnership in the UK. *Journal of Manipulative and Physiological Therapeutics* 25 (1), 21-33.

Lauridsen, H.H., Hartvigsen, J., Manniche, C., Korsholm, L. & Grunnet-Nilsson, N. (2006). Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BioMed Central Musculoskeletal Disorders* 7(82). doi: 10.1186/1471-2474-7-82.

Lawrence, D.J., Meeker, W., Branson, R., Bronfort, G., Cates, J.R., Haas, M. et al. (2008). Chiropractic Management of Low Back Pain and Low Back-Related Leg Complaints: A Literature Synthesis. *Journal of Manipulative and Physiological Therapeutics,* 31 (9), 659-673.

Leboeuf-Yde, C. (1999). Smoking and low back pain. A systematic literature review of 41 journal articles reporting 47 epidemiological studies. *Spine* 24 (14), 1463-1470.

Leboeuf-Yde, C. (2000). Alcohol and low back pain: a systematic literature review. *Journal of Manipulative and Physiological Therapeutics* 23 (5), 343-346.

Leboeuf-Yde, C., Andrén, J.A., Gernardt, M. & Malmqvist, S. (1997) Interprofessional contacts between chiropractors and other healthcare professionals in Sweden as seen from a chiropractic perspective. *Journal of Manipulative and Physiological Therapeutics* 20 (4), 241-245.

Leboeuf-Yde, C., Grønstvedt, A., Borge, J.A., Lothe, J., Magnesen, E., Nilsson, Ø. et al. (2004). The Nordic Back Pain Subpopulation Program: Demographic and Clinical Predictors for Outcome in Patients Receiving Chiropractic Treatment for Persistent Low-Back Pain. *Journal of Manipulative and Physiological Therapeutics* 27 (8), 493-502.

Leboeuf-Yde, C., Grønstvedt, A., Borge, J.A., Lothe, J., Magnesen, E., Nilsson, Ø. et al. (2005). The Nordic Back Pain Subpopulation Program: A 1-year prospective multicenter study of outcomes of persistent low-back pain in chiropractic patients. *Journal of Manipulative and Physiological Therapeutics* 28 (2), 90-96.

Leboeuf-Yde, C., Kyvik, K.O. & Bruun, N.H. (1998). Low back pain and lifestyle. Part I: Smoking. Information from a population-based sample of 29,424 twins. *Spine* 23 (20), 2207-2213.

Leboeuf-Yde, C., Nielsen, J., Kyvik, K.O., Feyer, R. & Hartvigsen, J. (2009) Pain in the lumbar, thoracic or cervical regions: do age and gender matter? A population-based study of 34,902 Danish twins 20-71 years of age. *Biomed Central Musculoskeletal Disorders*, 10, 39. http://www.biomedcentral.com/1471-2474/10/39

Leboeuf-Yde, C., Rosenbaum, A., Axén, I., Lövgren, P., Jørgensen, K., Halasz, L. et al. (2009) The Nordic Subpopulation Research Programme: prediction of treatment outcome in patients with low back pain treated by chiropractors-does the psychological profile matter? *Biomed Central Chiropractic and Osteopathy* 17, 14 doi:10.1186/1746-1340-17-14

Leboeuf-Yde, C., Yashin, A. & Lauritzen, T. (1996). Does smoking cause low back pain? Results from a population-based study. *Journal of Manipulative and Physiological Therapeutics* 19 (2), 99-108.

Leijon, O., Wahlström, J. & Mulder, M. (2009). Prevalence of Self-Reported Neck-Shoulder-Arm Pain and Concurrent Low Back Pain or Psychological Distress: Time-Trends in a General Population 1990-1996. *Spine* 34 (17), 1863-1868.

Lemieux, J., Beaton, D.E., Hogg-Johnson, S., Bordeleau, L.J. & Goodwin, P.J. (2007). Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *Journal of Clinical Epidemiology* 60, 448-455.

Licciardone, J.C. & Russo, D.P. (2006). Blinding Protocols, Treatment Credibility, and Expectancy: Methodologic Issues and Clinical Trials of Osteopathic Manipulative Treatment. *Journal of the American Osteopathic Association* 106 (8), 457-463.

Lindell, O., Johansson, S-E. & Strender, L-E. (2010). Predictors of stable return-to-work in non-acute, non-specific spinal pain: low total prior sick-listing, high self prediction and young age. A two-year prospective cohort study. *Biomed Central Family Practice* 11, 53 doi: 10.1186/1471-2296-11-53

Linton, S.J. (2000). A review of psychosocial risk factors in back and neck pain. *Spine* 25, 1148-1156.

Lønnberg, F., Pedersen, P.A. & Siersma, V. (2010). Early predictors of the long-term outcome of low back pain-results of a 22-year prospective cohort study from general practice. *Family Practice* 0, 1-6. doi: 10.1093/fampra/cmq061

McKee, M., Britton, A., Black, N., McPherson, K., Sanderson, C. & Bain, C. (1999). Interpreting the evidence: choosing between randomised and non-randomised studies. *British Medical Journal* 319, 312-315.

Macfarlane, G.J., Hunt, I.M. & Silman, A.J. (2000). Role of Mechanical and Psychosocial factors in the onset of forearm pain: prospective population based study. *British Medical Journal* 321, 1-5.

Mallen, C.D., Peat, G., Thomas, E., Dunn, K.M. & Croft, P.R. (2007). Prognostic factors for musculoskeletal pain in primary care: a systematic review. *British Journal of General Practice* 57, 655-661.

Malmqvist, S. (2001). *Psychometric properties of the Bournemouth Questionnaire in Swedish back pain patients: translation and validation*. Unpublished MSc Clinical Chiropractic Dissertation, University of Portsmouth, Portsmouth.

Malmqvist, S., Leboeuf-Yde, C., Ahola, T., Andersson, O., Ekström, K., Pekkarinen, H. et al. (2008). The Nordic back pain subpopulation program: predicting outcome among chiropractic patients in Finland. *Biomed Central Chiropractic & Osteopathy* 16, 13. doi: 10.1186/1746-1340-16-13

Maniadakis, N. & Gray, A. (2000). The economic burden of back pain in the UK. *Pain*, 84, 95-103.

Mannion, A.F., Porchet, F., Kleinstück, F.S., Lattig, F., Jeszensky, D., Bartanusz, V. et al. (2009). The quality of spine surgery from the patient's perspective: Part 2. Minimal clinically important difference for improvement and deterioration measured with the Core Outcome Measures Index. *European Spine Journal* 18 (suppl 3), S374-S379.

Martel, J., Dugas, C., Lalond, D. & Descarreaux, M. (2009). Validation of the French Version of the Bournemouth Questionnaire. *Journal of the Canadian Chiropractic Association* 53 (2), 102-110.

Maughan, E.F. & Lewis, J.S. (2010) Outcome measures in chronic low back pain. *European Spine Journal* 19 (9), 315-322. doi 10.1007/s00586-010-1353-6.

May, S., Gardiner, E., Young, S. & Klaber-Moffett, J. (2008). Predictor Variables for a Positive Long-Term Functional Outcome in Patients with Acute and Chronic Neck and Back Pain Treated with a McKenzie Approach: A Secondary Analysis. *The Journal of Manual and Manipulative Therapy* 16 (3), 155-160.

Menezes, C. L. da C., Henschke, N., Maher, C.G., Refshauge, K.M., Herbert, R.D., McAuley, J.H. et al. (2007). Prognosis of chronic low back pain: design of an inception cohort study. *BioMed Central Musculoskeletal Disorders* 8: 11. doi: 10.1186/1471-2474-8-11

Mercado, A.C., Carroll, L.J., Cassidy, J.D. & Côté, P.(2005). Passive coping as a risk factor for disabling neck or low back pain. *Pain* 117, 51-57.

Michaelson, P., Sjölander, P. & Johansson, H. (2004). Factors predicting pain reduction in chronic back and neck pain after multimodal treatment. *Clinical Journal of Pain* 20 (6), 447-454.

Moons, K.G.M., Altman, D.G., Vergouwe, Y & Royston P. (2009). Prognosis and prognostic research; application and improved clinical practice. *British Medical Journal* 338: b606. doi: 10.1136/bm.b606

Moons, K.G.M., Royston, P., Vergouwe, Y., Grobbee, D.E. & Altman, D.G. (2009). Prognosis and prognostic research: what, why, and how? *British Medical Journal* 338: b375. doi: 1136/bmj.b375

Mootz, R.D., Cherkin, D.C., Odegard, C.E., Eisenberg, D.M., Barassi, J.P. & Deyo, R.A. (2005). Characteristics of Chiropractic Practitioners, Patients, and Encounters in Massachusetts and Arizona. *Journal of Manipulative and Physiological Therapeutics,* 28 (9), 645-653.

Murphy, D.R., Hurwitz, E.L. & McGovern, E.E. (2009). A Nonsurgical Approach to the Management of Patients with Lumbar Radiculopathy Secondary to Herniated Disk: A Prospective Observational Cohort Study With Follow-Up. *Journal of Manipulative and Physiological Therapeutics,* 32 (9), 723-733.

Myers, S.S., Phillips, R.S., Davis, R.B., Cherkin, D.C., Legedza, A., Kaptchuck, T.J. et al. (2007). Patient Expectations as Predictors of Outcome in Patients with Acute Low Back Pain. *Journal of General Internal Medicine* 23 (2), 148-153.

Natvig, B., Ihlebaek, C., Grotle, M., Brage, S. & Bruunsgaard, D. (2010). Neck pain is often part of widespread pain and is associated with reduced functioning. *Spine* 35 (23), E1285-1289.

Ndetan, H.T., Bae, S., Evans, M.W. Jr, Rupert, R.L. & Singh, K.P. (2009). Characterization of health status and modifiable risk behaviour among United States adults using chiropractic care as compared with general medical care. *Journal of Manipulative and Physiological Therapeutics* 32 (6), 414-422.

Norman, G.R., Sloan, J.A. & Wyrwich, K.W. (2003). Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation. *Medical Care* 41, 582-492.

Ohnmeiss ,D.D. (2000). Repeatability of Pain Drawings in a Low Back Pain Population. *Spine* 25 (8), 980-988.

Ostello, R.W.J.G., Deyo, R.A., Stratford, P., Waddell, G., Croft, P., Von Korff, M. et al. (2008). Interpreting Change Scores for Pain and Functional Status in Low Back Pain: Towards International Consensus Regarding Minimal Important Change. *Spine* 33 (1), 90-94.

Ottenbacher, K.J., Johnson, M.B. & Hojem, M. (1988). The significance of clinical change and clinical change of significance: issues and methods. *The American Journal of Occupational Therapy* 42 (3), 156-163.

Ozegovic, D., Carroll, L.J., Cassidy, J.D. (2009). Does expecting mean achieving? The association between expecting to return to work and recovery in whiplash associated disorders: a population-based prospective cohort study. *European Spine Journal* 18, 893-899.

Parsons, S., Breen, A., Foster, N.E., Letley, L., Pincus, T., Vogel, S., & Underwood, M. (2007). Prevalence and comparative troublesomeness by age of musculoskeletal pain in different body locations. *Family Practice*, 24, 308-316.

Paterson, C. (1996). Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. *British Medical Journal* 312, 1016-1020.

Paterson, C. (2004). Seeking the patient's perspective: A qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Quality of Life Research* 13, 871-881.

Peat, G. (2004). PRA recommendations for low back pain-related functional limitation outcome measures CLEF 04. *The Chartered Society of Physiotherapy*. www.csp.org.uk Retrieved 18 August 2010.

Peat, G. (2008) Targeting treatment for non-specific musculoskeletal pain. *Pain* 139 (3), 483-384.

Phelan, S.M., van Ryn, M., Wall, M. & Burgess, D. (2009). Understanding Primary Care Physicians' Treatment of Chronic Low Back Pain: The Role of Physician and Practice Factors. *Pain Medicine* 10 (7), 1270-1279.

Pickard, A.S., Neary, M.P. & Cella, D. (2007). Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Quality Life Outcomes* 5: 70.

Pincus, T., Burton, A.K., Vogel, S. & Field, A.P. (2002). A Systematic Review of Psychological Factors as Predictors of Chronicity/Disability in Prospective Cohorts of Low Back Pain. *Spine* 27 (5), E109-120.

Pool, J.J.M., Ostelo, R.W.J.G., Knol, D., Bouter, L.M. & de Vet, H.C.W. (2010). Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain? *Manual Therapy* 15 (1), 111-116.

Pransky, G., Borkan, J.M., Young, A.E. & Cherkin, D. (2011) Are We Making Progress? The Tenth International Forum for Primary Care Research on Low Back Pain. *Spine* 36 (19), 608-1614.

Puhan, M.A., Frey, M., Buchi, S. & Schunemann, H.J. (2008). The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. *Health Quality Life Outcomes* 6: 46.

Quinn, J.V. & Wells, G.A. (1998). An assessment of clinical wound evaluation scales. *Academic Emergency Medicine* 5 (6), 583-586.

Rasmussen, C., Leboeuf-Yde, C., Hestbaek, L. & Manniche, C. (2008). Poor outcome in patients with spine-related leg or arm pain who are involved in compensation claims: a prospective study of patients in secondary care. *Scandinavian Journal of Rheumatology* 37 (6), 462-468.

Rejas, J., Pardo, A. & Ruiz, M.Á. (2008). Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patients-reported outcome measures. *Journal of Clinical Epidemiology* 61, 350-356.

Resnik, L., Liu, D., Hart, D.L. & Mor, V. (2008). Benchmarking Physical Therapy Clinic Performance: Statistical Methods to Enhance Internal Validity When Using Observational Data. *Physical Therapy* 88, 1078-1087.

Revicki, D.A. (2007). FDA draft guidance and health-outcomes research. *The Lancet* 369 (9561), 540-542.

Revicki, D., Hays, R.D., Cella, D. & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 61 (2), 102-109.

Roland, M & Morris, R. (1983). A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8, 141-144.

Rosen, M. (Chair). (1994) Epidemiology Review: the epidemiology and Cost of Back Pain. Clinical Standards Advisory Group HMSO, London 1994.

Rosenfeld, R.M. (1998). Meaningful Outcomes Research. In: Isenberg, S.F. (Ed) *Managed Care Outcomes, and Quality. A Practical Guide*, Chapter 9, pp. 99-115. New York: Thieme.

Rubinstein, S.M., Knol, D.L., Leboeuf-Yde, C., de Koekkoek, T.E., Pfeifle, C.E. & van Tulder, M.W. (2008). Predictors of a favourable Outcome in patients Treated by Chiropractors for Neck Pain. *Spine* 33 (13), 1451-1458.

Rubinstein, S.M., Leboeuf-Yde, C., Knol, D.L., de Koekkoek, T.E., Pfeifle, C.E. & van Tulder, M.W. (2007). The Benefits Outweigh the Risks for Patients Undergoing Chiropractic Care for Neck Pain: A Prospective, Multicenter, Cohort Study. *Journal of Manipulative and Physiological Therapeutics* 30 (6), 408-418.

Rubinstein, S.M., Leboeuf-Yde, C. Knol, D.L., de Koefkoek, T.E., Pfiefle, C.E. & van Tulder M.W. (2008). Predictors of Adverse Events Following Chiropractic Care For Patients With Neck Pain. *Journal of Manipulative and Physiological Therapeutics* 31, 94-103.

Samsa, G., Edelman, D., Rothman, M.L., Williams, G.R., Lipscomb, J. & Matchar, D. (1999). Determining clinically important differences in health status measures. *Pharmacoeconomics* 15 (2), 131-155.

Santilli, V., Beghi, E. & Finucci, S. (2006). Chiropractic manipulation in the treatment of acute back pain and sciatica with disc protrusion: a randomised double-blind clinical trial of active and simulated spinal manipulations. *The Spine Journal,* 6, 131-137.

Savingy, P., Kuntze, S., Watson, P., Underwood, M., Richie, G., Cotterell, M. et al. (2009) National Institute of Clinical Excellence. NICE clinical guideline 88 Low Back Pain: Early Management of Persistent non-specific low back pain. http://www.nice.org.uk/CG88quickrefguide

Schäfer, A., Hall, T., Müller, G. & Briffa, K. (2011). Outcomes differ between subgroups of patients with low back and leg pain following neural manual therapy: a prospective cohort study. *European Spine Journal* 20 (3), 482-490.

Schellingerhout, J.M., Verhagen, A.P., Heymans, M.W., Pool, J.J.M.P., Vonk, F., Koes, B.W. et al. (2008). Which subgroups of patients with non-specific neck pain are more likely to benefit from spinal manipulation therapy, physiotherapy, or usual care? *Pain* 139, 670-680.

Schellingerhout, J.M., Heymans, M.W., Verhagen, A.P., de Vet, H.C., Koes, B.W. & Terwee, C.B. (2011). Measurement properties of translated versions of neck-specific questionnaires:

a systematic review. *BMC Medical Research Methodology* 11, 87.
http://www.biomedcentral.com/1471-2288/11/87

Schers, H., Braspenning, J., Drijver, R., Wensing, M. & Grol, R. (2000) Low back pain in general practice: reported management and reasons for not adhering to the guidelines in the Netherlands. *British Journal of General Practice* 50 (457), 640-645.

Schers, H., Wensing, M., Huijsmans, Z., van Tulder & M., Grol, R. (2001) Implementation Barriers for General Practice Guidelines on Low Back Pain. A Qualitative Study. *Spine* 26, E348-E353.

Schmitt, M.A., De Wijer, A., van Genderen, F.R., van der Graaf, Y., Helders, P.J. & van Meeteren, N.L. (2009). The Neck Bournemouth Questionnaire cross-cultural adaptation into Dutch and evaluation of its psychometric properties in a population with subacute and chronic whiplash associated disorders. *Spine* 34 (23), 2551-2561.

Schulz, C.A., Hondras, M.A., Evans, R.L., Gudavalli, M.R., Long, C.R., Owens, E.F. et al. (2011). Chiropractic and self-care for back-related leg pain: design of a randomised clinical trial. *BioMed Central Chiropractic and Manual Therapies* 19: 8
http://chiromt.com/content/19/1/8

Scoggins, J.F. & Patrick, D.L. (2009). The use of patient-reported outcomes instruments in registered clinical trials: Evidence from ClinicalTrials.gov. *Contemporary Clinical Trials* 30, 289-292.

Scott, D., Brown, A., Lunt, I. & Thorne, L. (2004). Chapter 10: The Professions: Status and Qualifications. *Professional Doctorates: Integrating Professional and Academic Knowledge* pp. 139-149. England : Society for Research into Higher Education and Open University Press.

Shannon, H.S., Woodward, C.A., Cunningham, C.E., McIntosh, J., Lendrum, B., Brown, J. et al. (2001). Changes in general health and musculoskeletal outcomes in the workforce of a hospital undergoing rapid change: A longitudinal study. *Journal of Occupational Health Psychology* 6 (1), 3-14.

Skargren, E. & Öberg, B.E. (2008). Predictive factors for 1-year outcome of low-back and neck pain in patients treated in primary care: comparison between the treatment strategies chiropractic and physiotherapy. *Pain* 77, 201-207.

Skillgate, E. & Vingård, E., Josephson, M., Holm, L.W. & Alfredsson, L. (2009). Is smoking and alcohol consumption associated with long-term sick leave due to unspecific back or neck pain among employees in the public sector? Results of a three-year follow-up cohort study. *Journal of Rehabilitative Medicine* 41 (7), 550-556.

Sloan, J.A., Cella, D. & Hays, R.D. (2005). Clinical significance of patient-reported questionnaire data: another step toward consensus. *Journal of Clinical Epidemiology* 59 (12), 1217-1219.

Solberg, T.K., Sørlie, A., Sjaavik, K., Nygaard, Ø.P. & Ingebrigtsen, T. (2011). Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? *Acta Orthopaedica* 82 (1), 56-63.

Ssavedra-Hernádez, M, Castro-Sánchez, A.M., Fernández-de-las-Peñas, C., Cleland, J.A., Ortega-Santiago, R. & Arroyo-Morales, M. (2011). Predictors for Identifying Patients with Mechanical Neck Pain who are Likely to Achieve Short-Term Success with Manipulative Interventions Directed at the Cervical and Thoracic Spine. *Journal of Manipulative and Physiological Therapeutics* 34 (3), 145-152.

Staud, R. (2009). Chronic widespread pain and fibromyalgia: two sides of the same coin? *Current Rheumatology* 11 (6), 433-436.

Steenstra, I.A., Verbeek, J.H., Prinsze, F.J. & Knol, D.K. (2006). Changes in the incidence of occupational disability as a result of back and neck pain in the Netherlands. *Biomed Central Public Health* 6, 190-199.

Steenstra, I.A., Knol, D.L., Bongers, P.M., Anema, J.R., van Mechelen, W. & de Vet, H.C.W. (2009). What Works Best for Whom? An Exploratory, Subgroup Analysis in a Randomised, Controlled Trial on the Effectiveness of a Workplace Intervention in Low Back Pain Patients on Return to Work. *Spine* 34 (12), 1343-1349.

Terwee, C.B., Roorda, L.D., Dekker, J., Bierma-Zeinstra, S.M., Peat, G., Jordan, K.P. et al. (2010). Mind the MIC: large variation among populations and methods. *Journal of Clinical Epidemiology* 63, 524-534.

Testa, M.A. (1987). Interpreting Quality-of-Life Clinical Trial Data for Use in the Clinical Practice of antihypertensive therapy. *Journal of Hypertension* 5 (suppl 1), S9-S13.

Thiel, H.W. & Bolton, J.E. (2008) Predictors for immediate and global responses to chiropractic manipulation of the cervical spine. *Journal of Manipulative and Physiological Therapeutics* 31 (3), 172-183.

Troup, J.D.G. (1996) Back Pain and Epidemiology Review: The Epidemiology and Cost of Back Pain. *Social Science Medicine,* 42 (4), 561-563.

Tseng, Y-L., Wang, W.T.J., Chen, W-Y,, Hou, T.J., Chen, T-C. & Lieu, F-K. (2006). Predictors for immediate responders to cervical manipulation in patients with neck pain. *Manual Therapy* 11, 306-315.

Turk, D.C. (2000). Editorial, "Statisitical significance and clinical significance are not synonyms!" *The Clinical Journal of Pain* 16, 185-187.

Turk, D.C., Dworkin, R.H., Allen, R.R., Bellamy, N., Brandenberg, N., Carr, D.B. et al. (2003) Core Outcome Domains for Chronic Pain in Clinical Trials: IMMPACT recommendations. *Pain* 106, 337-345.

Turner, D., Schünemann, H.J., Griffith, L.E., Beaton, D.E., Griffiths, A.M., Critch, J.N. et al. (2009). Using the entire cohort in the receiver operating characteristic analysis maximises precision of the minimal important difference. *Journal of Clinical Epidemiology* 62, 374-379.

Turner, D., Schünemann, H.J., Griffith, L.E., Beaton, D.E., Griffiths, A.M., Critch, J.N. et al. (2010). The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology* 63, 28-36.

Udén, A., Aström, M. & Bergenudd, H. (1988). Pain drawings in chronic back pain. *Spine* 13 (4), 389-392.

Underwood, M.R., Morton, V. & Farrin, A. (2007). Do baseline characteristics predict response to treatment for low back pain? Secondary analysis of the UK NEAM dataset [ISRCTN32683578]. Rheumatology 46, 1297-1302.

Van Middelkoop, M., Rubinstein, S.M., Kuijpers, T., Verhagen, A.P., Ostelo, R., Koes, B.W., et al. (2011) A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. *European Spine Journal* 20, 19-39.

Van Tulder, M., Becker, A., Bekkering, T., Breen, A., Carter, T., Del Real, M.T.G., Hutchinson, A., Koes, B. et al. (2004). European Guidelines for the Management of Acute Non-Specific Low Back Pain in Primary Care. *European Commission Research Directorate General*. http://www.backpaineurope.org/web/html/wg1_results.html

Verbeek, J., Sengers, M.J., Riemens, L. & Haafkens, J. (2004) Patient expectations of treatment for back pain: a systematic review of qualitative and quantitative studies. *Spine* 29 (20), 2309-2318.

Vernon, H. & Mior, S. (1991). The Neck Disability Index: A Study of Reliability and Validity. *Journal of Manipulative and Physiological Therapeutics* 14 (7), 409-415.

Voorhies, R.M., Jiang, X. & Thomas, N. (2007). Predicting outcome in the surgical treatment of lumbar radiculopathy using the Pain Drawing Score, McGill Short Form Pain Questionnaire, and risk factors including psychosocial issues and axial joint pain. *Spine* (5), 516-524.

Waddell, G. (1987) A new model for the treatment of low back pain. *Spine* 12 (7), 632-644.

Waddell, G. (2005). Subgroups within non-specific low back pain. *The Journal of Rheumatology* 32 (3), 395-396.

Waddell, G., McIntosh, A., Hutchinson, A., Feder, G & Lewis, M. (1999). Low Back Pain Evidence Review. London. Royal College of General Practitioners. http://www.rcgp.org.uk

Wand, B.M., McAuley, J.M., Marston, L. & De Souza, L.H. (2009). Predicting Outcome in Acute Low Back Pain Using Different Models of Patient Profiling. *Spine* 34(18), 1970-1975.

Webb, R., Brammah, T., Lunt, M., Unwin, M., Allison, T., & Symmons, D. (2003). Prevalence and Predictors of Intense, Chronic and Disabling Neck Pain in the UK General Population. *Spine*, 28 (11), 1195-1202.

Wells, G.A., Tugwell, P., Kraag, G., Baker, P.R.A., Groh, J. & Redelmeier, D.A. (1993). Minimum important difference between patients with Rheumatoid Arthritis: the patient's perspective. *Journal of Rheumatology* 20 (3), 557-560.

Werneke, M.W. & Hart, D.L. (2004) Categorizing Patients with Occupational Low Back Pain by Use of the Quebec Task Force Classification System versus Pain Pattern Classification Procedures: Discriminant and Predictive Validity. *Physical Therapy* 84 (3), 243-254.

Wyatt, M., Underwood, M.R., Scheel, I.B., Cassidy, J.D. & Nagel, P. (2004). Back pain and health policy research: the what, why, how, who and when. *Spine* 29, E468-475.

Wyrwich, K.W. & Wolinsky, F.D. (2000). Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice* 6 (1), 39-49.

Zeiger, R.I. (2004). Towards Continuous Medical Education. *Journal of General Internal Medicine* 20, 91-94.

Zvolensky, M.J., McMillan, K., Gonzalez, A. & Asmundson, G.J. (2009). Chronic pain and cigarette smoking and nicotine dependence among a representative sample of adults. *Nicotene Tobacco Research* 11 (12), 1407-1414.

Appendix i

New Patient Questionnaire Including the Pain diagram

Appendix ii

Pre-treatment BQ

Appendix iii

Grid for Assessing Pain Diagram
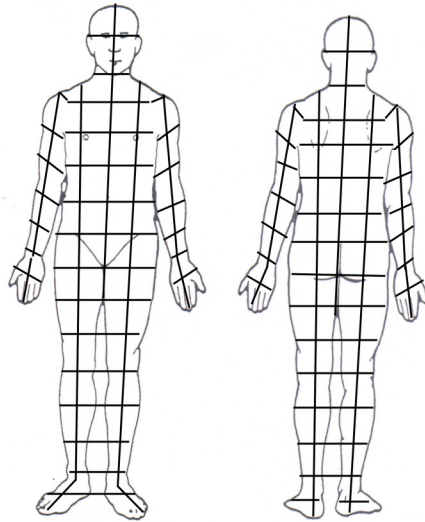
Appendix iv

Post Treatment BQ with Patient Global Improvement Scale

Appendix v

Copies of letters from the Anglo-European College of Chiropractic Ethics Sub-committee

Table A. Crude Odds Ratios for Baseline Variables with All Three Methods of Assessing Improvement in Back Pain Patients at treatment 4/5 and treatment 10 (significant interactions in red)

Table A. Crude Odds Ratios of Baseline Variables with Three Methods of Categorising Improvement at treatment 4/5 (1) and 10 (2)

	Direct 1			Direct 2			Anchor 1			Anchor 2			Distribution 1			Distribution 2		
Variable	OR	95% CI	P sig<0.05	OR	95% CI	P sig<0.05	OR	95% CI	P sig<0.05	OR	95% CI	P sig<0.05	OR	95% CI	P sig <0.05	OR	95%CI	P sig<0.05
Gender (M)	1.12	0.93 to 1.35	0.22	0.94	0.56 to 1.58	0.82	0.91	0.76 to 1.1	0.33	1.31	0.71 to 1.8	0.6	0.94	0.77 to 1.16	0.58	0.82	0.47 to 1.44	0.49
Inappropriate Signs	0.80	0.61 to 1.05	0.1	0.49	0.24 to 1.0	0.05	0.94	0.71 to 1.23	0.63	0.59	0.30 to 1.19	0.14	0.81	0.61 to 1.09	0.16	0.43	0.20 to 0.91	0.26
Radiating pain	1.04	0.86 to 1.26	0.72	0.99	0.58 to 1.68	0.93	1.22	1.01 to 1.48	0.04	2.01	1.24 to 3.24	0<0.01	1.17	0.95 to 1.45	0.14	1.10	0.62 to 1.96	0.74
Widespread pain	0.64	0.43 to 0.95	0.03	0.37	0.15 to 0.91	0.03	1.04	0.69 to 1.56	0.85	1.11	0.45 to 2.71	0.83	1.22	0.77 to 1.93	0.41	2.70	0.60 to 11.78	0.2
Similar pain in past yes	1.04	0.85 to 1.27	0.73	0.91	0.51 to 1.62	0.74	1.17	0.96 to 1.43	0.12	1.09	0.66 to 1.83	0.73	1.19	0.96 to 1.48	0.12	1.16	0.63 to 2.15	0.64
Chronic <7 weeks	3.10	2.56 to 3.77	0<0.01	4.39	2.45 to 7.88	0<0.01	2.70	2.23 to 3.27	0<0.01	3.05	1.89 to 4.94	0<0.01	2.73	2.22 to 3.37	0<0.01	3.56	1.90 to 6.67	0<0.01
Medication use	1.12	0.90 to 1.39	0.32	0.74	0.42 to 1.3	0.3	1.50	1.20 to 1.87	0<0.01	1.54	0.91 to 2.62	0.11	1.40	1.09 to 1.8	0.01	1.23	0.60 to 2.35	0.54
Job Satisfaction	1.32	0.94 to 1.85	0.11	0.91	0.35 to 2.4	0.85	1.07	0.76 to 1.49	0.71	0.70	0.29 to 1.66	0.42	1.05	0.72 to 1.52	0.81	0.33	0.75 to 1.45	0.14
Work Normally	1.42	0.98 to 2.06	0.06	3.940	1.49 to 10.46	0.01	1.55	1.07 to 2.24	0.02	1.70	0.65 to 4.46	0.28	1.43	0.97 to 2.11	0.08	1.69	0.57 to 4.99	0.34
Smoker	1.09	0.91 to 1.31	0.37	1.57	0.93 to 2.65	0.09	1.18	0.98 to 1.41	0.09	1.55	0.97 to 2.47	0.07	1.27	1.04 to 1.55	0.02	1.45	0.82 to 2.56	0.20
Alcohol	0.92	0.75 to 1.14	0.46	1.07	0.60 to 1.92	0.81	0.87	0.70 to 1.07	0.19	0.61	0.36 to 1.05	0.72	0.89	0.71 to 1.13	0.36	1.18	0.63 to 2.21	0.60
Phys Activity	1.38	1.09 to 1.74	0.01	1.71	0.92 to 3.2	0.09	1.04	0.82 to 1.31	0.77	0.95	.53 to 1.71	0.86	0.96	0.74 to 1.25	0.78	0.78	0.37 to 1.64	0.51
General Health	1.81	1.38 to 2.37	0<0.01	2.29	1.214.32	0.01	1.24	0.95 to 1.62	0.11	0.80	0.43 to 1.48	0.48	1.20	0.90 to 1.6	0.21	0.77	0.35 to 1.68	0.51
Pain Diagram<8	1.31	1.09 to 1.58	0<0.01	1.80	1.07 to 3.04	0.03	1.14	0.94 to 1.37	0.18	1.36	0.85 to 2.16	0.20	1.12	0.92 to 1.38	0.25	1.12	0.6 to 1.98	0.69
Employed	1.17	0.63 to 2.17	0.63	8.80	1.66 to 46.79	0.01	1.26	0.67 to 2.33	0.47	1.07	0.23 to 4.88	0.94	1.11	0.56 to 2.19	0.76	0.84	0.10 to 7.21	0.88
Treatment Expectation	2.46	2.02 to 2.99	0<0.01	2.15	1.27 to 3.65	0.01	2.06	1.69 to 2.5	0<0.01	2.94	1.77 to 4.8	0<0.01	2.02	1.64 to 2.49	0<0.01	2.54	1.43 to 4.51	0<0.01
Age <50.1	1.07	0.86 to 1.32	0.52	1.04	0.58 to 1.85	0.90	1.35	1.09 to 1.67	0.01	1.09	0.65 to 1.82	0.74	1.30	1.04 to 1.6	0.02	2.16	1.19 to 3.9	0.01
Trauma	1.11	0.89 to 1.37	0.35	1.24	0.67 to 2.31	0.49	1.05	0.83 to 1.34	0.68	1.41	0.82 to 2.42	0.22	1.10	0.87 to 1.39	0.43	1.14	0.58 to 2.22	0.70
Tot BQ<30	0.63	0.52 to 0.76	0<0.01	0.53	0.31 to 0.9	0.02	0.15	0.12 to 0.18	0<0.01	0.12	0.07 to 0.21	0<0.01	0.18	0.15 to 0.23	0<0.01	0.23	0.12 to 0.42	0<0.01

Table B. Crude Odds Ratios for Baseline Variables with the Anchor-based Method of Assessing Improvement in Neck Pain Patients at treatment 4/5 and treatment 10 (significant interactions in red)

Table B. Crude Odds Ratios of Baseline Variables with the Anchor-Based Method of Assessing Improvement at Treatment 4/5 (1) and Treatment 10 (2)

	Anchor 1			Anchor 2		
Variable	OR	95% CI	P(sig<0.05)	OR	95% CI	P (sig<0.05)
Gender (M)	0.96	0.76 to 1.22	0.75	1.05	0.54 to 2.04	0.90
Inappropriate Signs	1<0.01	0.75 to 1.34	0.99	1.05	0.48 to 2.32	0.90
Radiating Pain	1.11	0.87 to 1.42	**0.41**	1.21	0.62 to 2.36	**0.58**
Widespread pain	1.14	0.73 to 1.77	0.57	5.36	0.66 to 43.45	0.12
Similar pain	0.90	0.70 to 1.16	0.41	0.95	0.47 to 1.91	0.89
Chronic <7 weeks	2.20	1.72 to 2.8	0<0.01	3.07	1.50 to 6.29	0<0.01
Medication use	1.59	1.21 to 2.08	0<0.01	0.85	0.42 to 1.72	**0.66**
Job Satisfaction	0.92	0.63 to 1.34	0.65	0.83	0.29 to 2.31	0.71
Work Normally	1.56	0.95 to 2.56	0.08	2.33	0.74 to 7.35	0.15
Smoker	1.23	0.97 to 1.55	0.09	2.07	1.07 to 3.1	0.03
Alcohol	0.87	0.67 to 1.12	0.28	0.73	0.36 to 1.49	0.39
Physical Activity	1.02	0.76 to 1.35	0.92	0.64	0.29 to 1.41	0.27
General Health	1.26	0.92 to 1.74	0.15	1.20	0.52 to 2.77	0.68
Pain Diagram <8	1.01	0.80 to 1.28	0.91	0.84	0.44 to 1.61	0.60
Employed	1.44	1.07 to 1.94	0.02	1.42	0.65 to 3.11	0.38
Treatment Expectation	1.52	1.19 to 1.95	0<0.01	2.75	1.41 to 5.37	0<0.01
Age <50.1	1.26	0.96 to 1.66	0.10	1.40	0.65 to 2.98	0.39
Trauma	1.86	0.80 to 4.29	0.15	1.52	1.86 to 4.29	0.15
Total BQ<30	0.15	0.11 to 0.19	0<0.01	0.36	0.18 to 0.7	0<0.01

Appendix viii

Chiropractic & Manual Therapies 2011, 19:27.