The development, evaluation, and initial implementation of a national programme for the use and collation of patient reported outcome measures (PROMs) in osteopathic back pain services in the UK.

Carol Fawkes

Centre for Primary Care and Public Health

Barts and The London Queen Mary's School of Medicine and Dentistry

Thesis submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

2016

Abstract

Introduction

The use of Patient Reported Outcome Measures (PROMs) to measure the effects of care is being advocated increasingly in clinical settings. Current patient data capture involves completion of paper questionnaires which is costly and environmentally perplexing. New innovations are required to balance the challenges of introducing data capture directly from patients while considering budgets, access to Information Technology, and the capability to use technological devices.

Methods

Two qualitative studies were undertaken to identify the views of patients and clinicians concerning electronic PROM data capture in osteopathic practice. One qualitative study involved patient interviews to identify their views on a selection of specific PROMs. Clinician focus groups and interviews (osteopaths, chiropractors, and physiotherapists) were undertaken concerning their views and experiences of using PROMs. Scoping of PROMs in musculoskeletal practice was undertaken followed by a systematic review of one identified PROM. The review and qualitative work informed the development of content for a mobile and web app for capturing PROM data. The app was piloted to evaluate feasibility, and the clinimetric performance of the included PROMs. Feedback from the pilot informed revisions to the app prior to implementation into osteopathic practice.

Results

Clinicians (n=46) identified a range of barriers and facilitators to PROM use. Patients (n=22) while generally more enthusiastic than clinicians welcomed the opportunity to provide feedback and although undaunted by the use of technology highlighted the need for assurances concerning confidentiality of data, and limits on data sharing. The systematic review identified good measurement properties for the Bournemouth (BQ). Piloting of the app involved 257 participants contributing 404 data returns: it performed well requiring minimal revision prior to implementation.

Conclusions

The app performed well demonstrating great potential for further development to collect outcome data in a musculoskeletal clinical setting.

Summary of Contents

Contents

Appendices

List of figures

List of tables

Acknowledgements

I would like to express my sincere thanks to my supervisors Dr Dawn Carnes and Dr Rob Froud for their guidance throughout the PhD process. I would also like to thank members of the National Council for Osteopathic Research who have supported development of the PROM app, and the Institute of Osteopathy for funding the software development. I am grateful to Jonathan Field for sharing his experiences of developing the Care Response system. I am indebted to the osteopaths, chiropractors and physiotherapists who gave their time in the focus groups and interviews, and the osteopaths who participated in the pilot phase of the project. Finally, and most importantly, I would like to thank the patients who participated in interviews, and who have used the app during the pilot process – without their input the project would have lacked an essential element.

1

Introduction

More than 150 years ago, Florence Nightingale reported that patients left her care either "dead, relieved, or unrelieved" (Nightingale, 1858). In spite of the intervening passage of time, the National Health Service (NHS) has continued to collect routine data concerning whether patients leave hospital dead or alive (HSCIC, 2015), but has focussed few resources on assessment of delivery of care and its quality. Lord Darzi, in his evidence to the parliamentary health committee in 2008, highlighted that mortality statistics were an imprecise and perhaps inappropriate statistic to measure success in many conditions (Darzi, 2008; Howie and Hamilton, 2013). In his publication "High Quality Care For All – NHS Next Stage Review " Lord Darzi focussed on the need for quality in healthcare, and patient-focussed care (Darzi, 2008). Ongoing quality assessment is an implicit part of good patient care, and this can be achieved in many ways. The manner in which the quality of patient care is assessed depends as much on clinically-led initiatives as on political changes introduced.

1.1.1 Historical context

Historically quality assessment has been through monitoring of scientific measures *e.g.* clinical tests; clinical audit; and ongoing data collection. More recent changes have focussed on the introduction of patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). Nationally and internationally there is an increasing focus in healthcare and other service providers on collection of outcome data. In some instances the level of data collection and retention is both overwhelming and superfluous; for some individuals this widespread data collection

and its distribution is a cause for concern especially when related to personal information. While many of the issues highlighted have focussed on the NHS, the drive to measure outcomes of care is no less relevant to osteopathic practice. The growth of an evidence based culture in health care has been evident in the past decade. Commissioners of health *e.g.* Clinical Commissioning Groups (CCGs), and private health insurers increasingly require evidence before considering the purchase of services including osteopathy. Meeting the demands of an evidence-based culture represents both challenges and opportunities for osteopaths and their patients. The focus for this study will be the introduction of PROMs as part of day-to-day data collection into osteopathic practice in the United Kingdom (UK).

1.1.2 Aims and objectives

I had two main aims for conducting this research, they were:

I. To design and develop a nationwide system of collecting routine PROM data from osteopathic patients;
II. To enable the establishment of baseline standards for outcomes for patients presenting with musculoskeletal conditions to osteopaths as a benchmark comparator.

In order to meet the aims of the study, a series of research objectives were identified and undertaken including:

i. Undertaking a review of the literature concerning the use of PROMs in clinical practice, and the different formats in which they have been used;
ii. Conducting qualitative work to identify the views of patients about the concept of data collection in clinical practice, and their views concerning three different patient reported outcome measures (PROMs);
iii. Conducting qualitative work to identify the views of osteopaths on the concepts of data collection in clinical practice;
iv. Conducting qualitative work to identify the views (and experiences) of physiotherapists and chiropractors concerning PROMs and their use in a clinical setting.

v. Scoping, and systematically reviewing the literature concerning the measurement properties of a selection of key PROMs;
vi. Using the review and qualitative findings to develop content for an app suitable for use via the Internet, mobile telephone or other mobile device *e.g*. a Tablet computer;
vii. Pilot testing the app to assess its functionality, the feasibility of using the app in clinical practice, and the clinimetric performance of the PROMs in an electronic format;
viii. Examination of the responsiveness of the PROMs in UK osteopathic clinical settings and identify baseline standards for patients attending with musculoskeletal symptoms;
ix. Examination of the test-retest reliability of the PROMs in UK osteopathic clinical settings and identify baseline standards for patients attending with musculoskeletal symptoms;
x. Examination of data concerning patient satisfaction and experience in clinical practice;
xi. Refinement of the app based on feedback, and its implementation into day-to-day osteopathic practice.

To begin to undertake this PhD, an initial literature search was conducted to identify existing literature, and where gaps existed, or existing studies could be developed more fully. The strategy for the literature review is shown in Figure 1.1, and the bibliographic framework to underpin the literature search is shown in Figure 1.2.

The complete outline to this thesis will be described in section 1.3 at the end of this chapter.

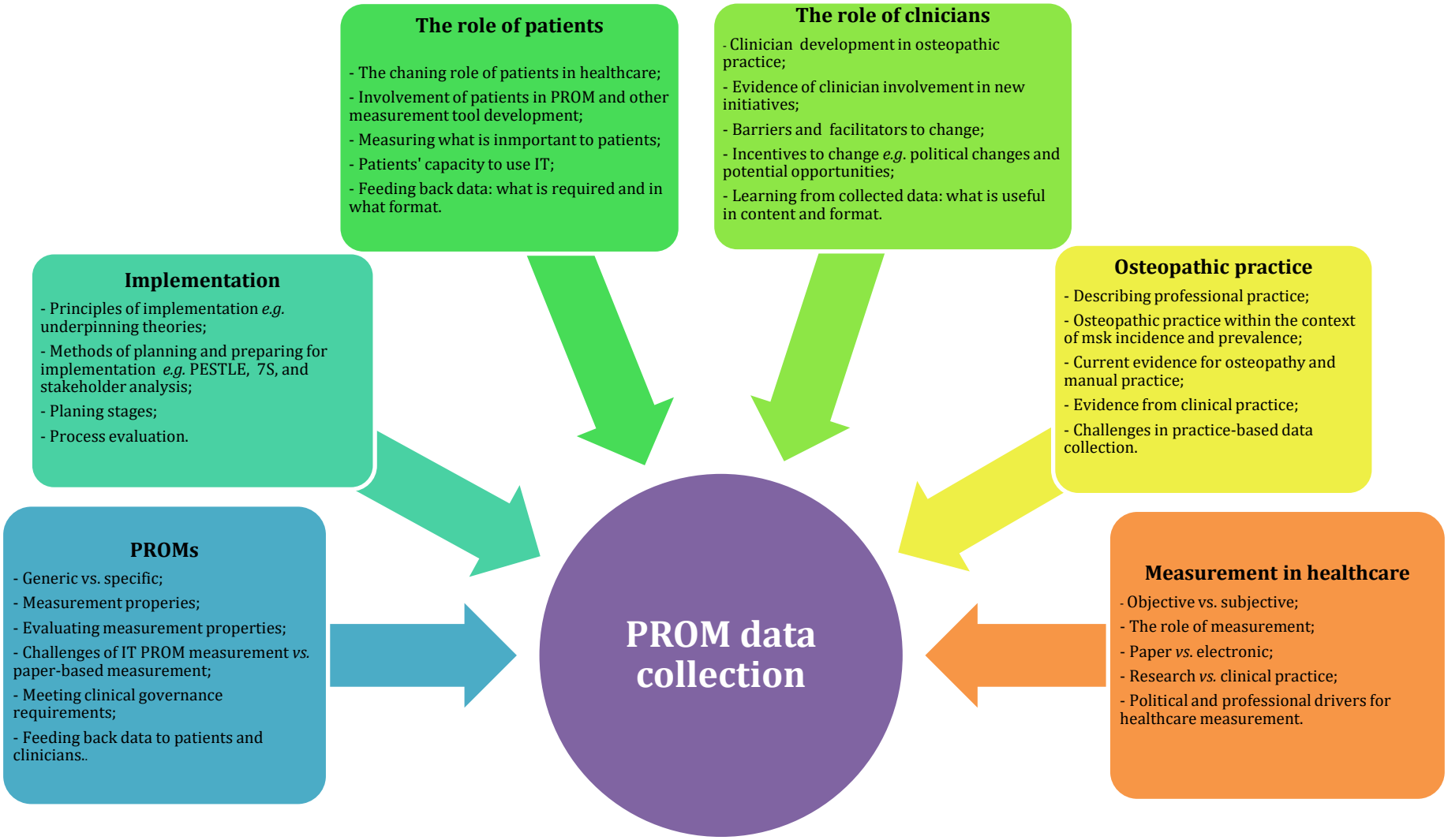


Figure 1.1 Strategy for the literature review

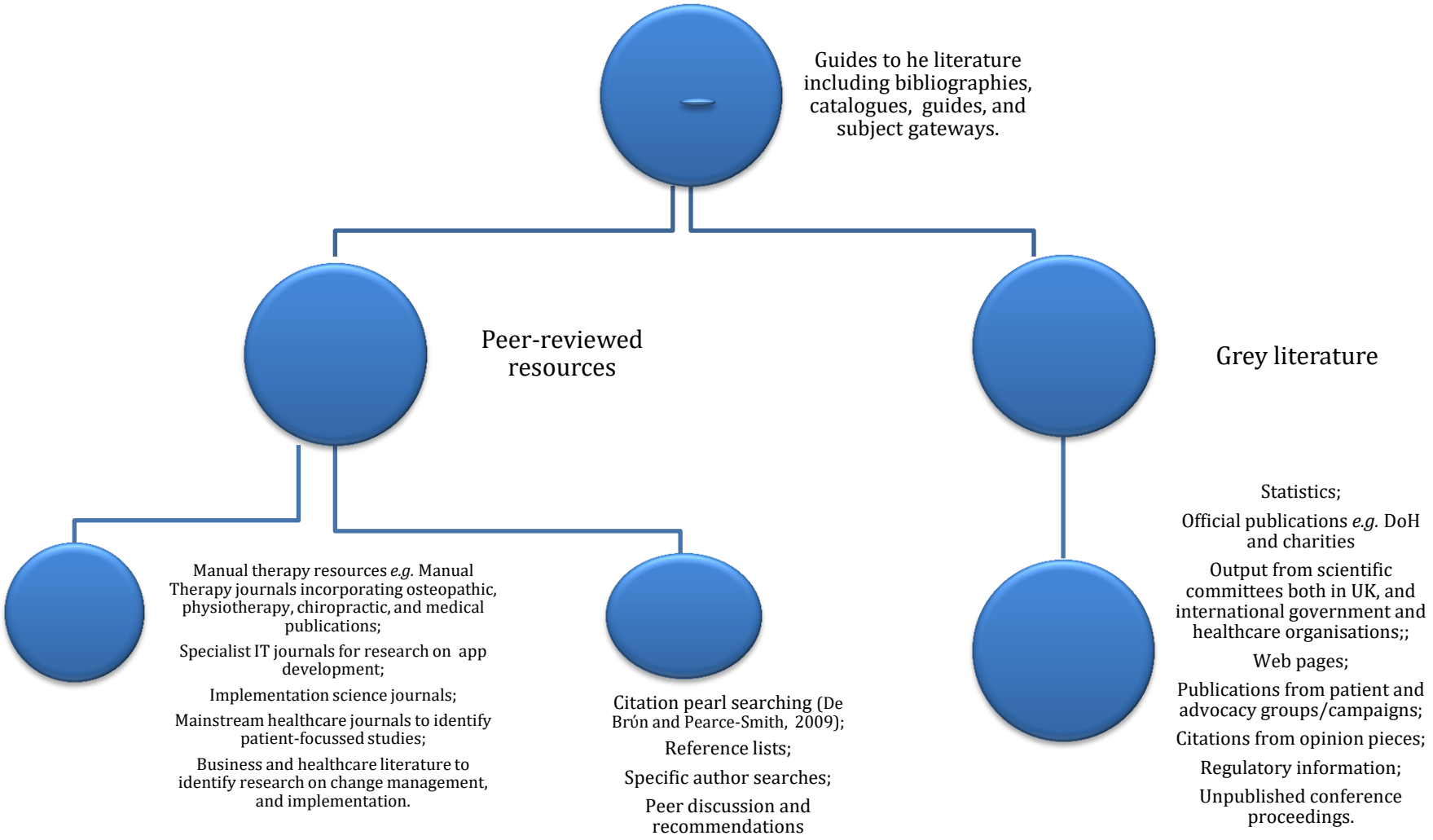


Figure 1.2 Bibliographic framework to underpin the literature search (Hart, 2003)

1.2 Background

1.2.1 The changing face of osteopathic practice

Osteopathic medicine which incorporates what is recognised as "osteopathy outside the USA" was founded in 1874 in Kirksville, Missouri, by Andrew Taylor Still. It became a recognised diagnostic and therapeutic system independent of the established medical thinking of the time. Still trained formally as a doctor but sought new methods of treatment emphasising physical treatment for systemic diseases, not just musculoskeletal conditions. His intent was not to create a separate profession but to give the current system of medicine a more rational and scientific basis. Still organised a school in Kirksville, Missouri, for the teaching of osteopathy; this approach finally received recognition in all 50 states in the USA in 1973. Osteopaths in the USA are awarded the Diploma in Osteopathy (DO). The first school of osteopathy was set up in the United Kingdom (UK) in London in 1917 by John Martin Littlejohn. Osteopathy developed in the UK, but was an unregulated profession. Osteopaths who had received training at accredited colleges of osteopathy were awarded a DO joined a voluntary register, the General Council and Register of Osteopaths (GCRO) and were entitled to use the post nominal letters DO; MRO (Member of the Register of Osteopaths).

The passing of the Osteopaths' Act (1993) introduced many changes. The profession became regulated by statute under the aegis of the General Osteopathic Council (GOsC). The title of 'Osteopath' became protected in law; training and practise became regulated and, in May 2004, mandatory continual professional development (CPD) was introduced bringing osteopathy into line with other healthcare professions. Training in osteopathy in the UK has taken the form of a four year BSc degree since 1990; MSc degrees are increasingly awarded by Osteopathic Educational Institutions (OEIs) in the UK as their education programmes have developed.

It was accepted practice for osteopaths prior to 1989 to be allowed two advertisements after qualification announcing they were in practice. Once that had occurred their practices were developed through word-of-mouth, outreach visits to local groups, and personal networks. The Office of Fair Trading (OFT) published a

paper in 1986 to lift restrictions on advertising for osteopaths. Despite objection from the GCRO, the restriction was lifted by the Monopolies and Mergers Commission in 1989 (Monopolies and Mergers Commission, 1989; Ogus, 2004). At that time the OFT was responsible for the protection of consumer interests, and this role has now passed to a range of organisations including the Advertising Standards Authority (ASA). The ASA is described as "the UK's independent regulator of advertising across all media", and on its web site home page it describes its work thus:

"We apply the Advertising Codes, which are written by the Committees of Advertising Practice. Our work includes acting on complaints and proactively checking the media to take action against misleading, harmful or offensive advertisements. Our ambition is to make every UK ad a responsible ad." (ASA, 2016a).

While at the time the pronouncement from the Monopolies and Mergers Commission seemed to be a beneficial initiative, the ability to advertise osteopathic services has not been without its challenges. In Rule 12.1, the ASA is quite clear on its requirements for underpinning evidence to support claims for advertising services in relation to the delivery of healthcare interventions. It states:

"Objective claims must be backed by evidence, if relevant consisting of trials conducted on people. Substantiation will be assessed on the basis of the available scientific knowledge" (ASA Relevant Code Rule 12.1, 2016b*).*

Unfortunately, an emphasis on evidence can all too often be focusses on proving the benefits of practice instead of <u>improving</u> practice. One of the key drivers for the introduction of Patient Reported Outcome Measures (PROMs) and Patient Reported Experience Measures (PREMs) into practice has been to enhance quality of care: meeting the demands of managers and commissioners while giving patients the opportunity to engage and feedback about their management (NHS England, 2015a). While the introduction of PROMs and PREMs is helpful in collecting data concerning the outcome of care, the nature of what constitutes "evidence" is still debated by some commentators.

1.2.2 The challenge of evidence

It has been said that the nature of what constitutes "evidence" and "available scientific knowledge" lacks uniformity. This highlights a debate concerning the nature of evidence that has been ongoing for some time. In 1967, Feinstein's publication *Clinical Judgement* highlighted flaws in the traditional approach to medical decision-making based on bias and the process of clinical reasoning (Feinstein, 1967). Cochrane highlighted the lack of clinical trials underpinning interventions, and Eddy described gaps in the evidence for patient management (Cochrane, 1972; Eddy, 1982a; Eddy, 1982b; Eddy, 1984; Eddy, 1988). This led to the development of the term "evidence-based" by Eddy, and its use in policy and practice guideline development (Eddy, 1990). In 1992, Guyatt and colleagues published the term "evidence-based medicine" (EBM) within the context of medical education (Evidence-based medicine working group, 1992). Sackett *et al.* expanded upon the term EBM to explain what it did and did not include (Sackett *et al.,* 1996). While the value of evidence-based medicine is much debated by clinicians, it remains in common parlance. More recent commentators prefer the term "evidence-informed medicine/practice" highlighting the role of clinicians' experiences working in a coordinated manner with best available evidence (Nevo and Slonim-Nevo, 2011; Smith and Rennie, 2014).

While agreeing on appropriate terminology is important, it is equally important to agree what constitutes evidence. In 1979, the Canadian Task Force on the Periodic Health Examination were one of the first groups to popularise the application of a hierarchy of evidence. In the intervening years a range of different hierarchies have been developed but the Canadian version is shown in Figure 1.3.

While the systematic review and meta-analysis are widely regarded as useful tools to consolidate vast amounts of knowledge into manageable forms for busy clinicians, they are not without their critics. Moore and Jull (2006) noted that a single systematic review is only as useful as the quality of included trials, and some of those trials may not reflect current standards of practice, or assess the quality of the interventional approach. In further developments, the systematic review of systematic reviews has begun to appear in the literature. While this is regarded by

some as "an untrusted methodology at best" (Moore and Jull, 2006), it could be regarded as taking clinicians even further away from primary data than a systematic review. The value of systematic reviews will be discussed in greater length in Chapter 4.

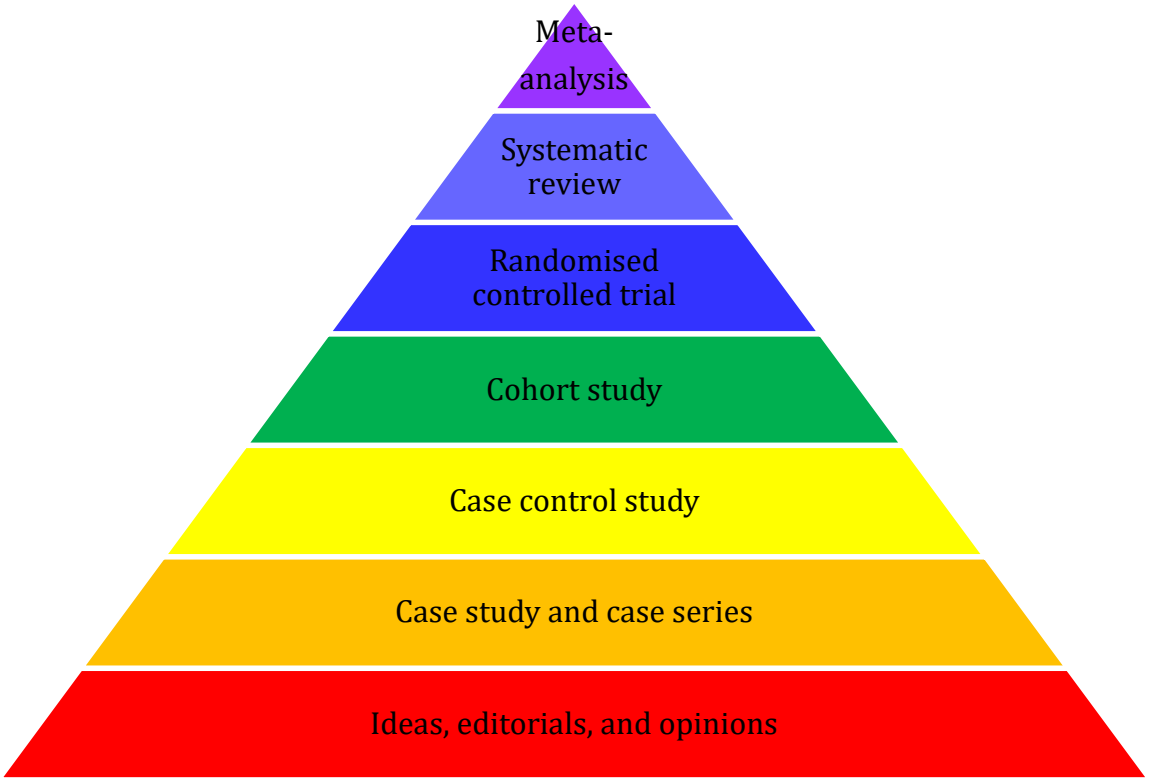


Figure 1.3. The Hierarchy of evidence (adapted from the Canadian Task Force on the Periodic Health Examination, 1979).

Despite the misgivings of some commentators concerning systematic reviews, meta-analyses, and systematic reviews of systematic reviews, the randomised controlled trial remains for many the "gold standard" in research (Evans, 2003). Jones and Podolsky note:

"*When the 1962 amendments to the US Food and Drug Administration mandated proof of efficacy through "well-controlled" studies—namely, RCTs—before new drug approval, the US Government set the stage for the avalanche of pharmaceutical trials that followed*" (Jones and Podolsky, 2015).

The very concept of a "gold standard" is challenging for some authors. This term, first used in 1962 and inspired by the Olympic meaning of the best available, has been increasingly used in relation to research (Anonymous, 1962; Rudd, 1979). The idea of reaching perfection with such a standard was regarded by some scientists as presumptuous since knowledge should be evolving continually (Duggan, 1992): for others the concept of a gold standard indicated the best tool available at the time (Versi, 1992; Claassen, 2005). The debate concerning the appropriateness of hierarchies, and the place of the RCT therein, continued in the Harveian Oration delivered by Sir Michael Rawlins. He stated

"*The notion that evidence can be reliably placed in hierarchies is illusory. Hierarchies place RCTs on an uncomfortable pedestal for while the technique has advantages it also has disadvantages. Observational studies have defects but they also have merit. Decision makers need to assess and appraise all the available evidence irrespective as to whether it has been derived from RCTs or observational studies; and the strengths and weaknesses of each need to be understood if reasonable and reliable conclusions are to be drawn. Nor, in reaching these conclusions, is there any shame in accepting that judgements are required about the 'fitness for purpose' of the components of the evidence base"* (Rawlins, 2008).

Jadad and Enkin expand upon the advantages and disadvantages cited (Jadad and Enkin, 2007). The place of observational and cohort designs within the hierarchy has been debated. While they are regarded by some authors as being at greater risk of systematic error than RCTs (Chalmers *et al.,* 1983; Miller, 1989), later comparison studies of the results of studies using each design have questioned the strength of this claim (Benson and Hartz, 2000; Concato *et al.,* 2000; Evans, 2003). For other commentators, however, observational studies have the added value of being able to use larger and more diverse populations with common comorbidities (Sørensen *et al.,* 2006; Silverman, 2009).

While the use of research designs other than RCTs is growing in acceptance, the use of day-to-day data capture has been introduced increasingly into clinical practice. Such data capture has been achieved through the use of Patient Reported Outcome Measurements (PROMs) and Patient Reported Experience Measures (PREMs).

1.2.3 National PROMs initiatives

Since 2009, Patient Reported Outcome Measurement (PROM) data have been collected and published through four national PROMs programmes in NHS surgical care for hip and knee arthroplasty, varicose vein surgery, and inguinal hernia repair (Standard NHS Contracts, 2008; National PROMs Programme, 2009; HES, 2014; HSCIC, 2015).

Although current PROMs data collection focusses on four specific areas offering remedies to distressing symptoms, the reports of other musculoskeletal symptoms in the population remain high. Low back pain and other musculoskeletal conditions are widespread in their incidence and prevalence. It is estimated that 37% of individuals in Britain will experience back pain every year which has lasted more than one day (OPCS, 1994). In its 2009 development of guidelines for chronic non-specific low back pain, The National Institute for Health and Care Excellence (NICE) gave a more up-to-date summary stating "Estimates of the prevalence of low back pain vary considerably between studies - up to 33% for point prevalence, 65% for 1- year prevalence, and 84% for lifetime prevalence (Walker *et al.,* 2000)" (Savigny *et al.,* 2009). Around 20% of these patients (1 in 15 of the population) will consult their GP about their back pain (Macfarlane *et al.,* 2006). This results in 2.6 million people, in the UK, seeking advice about back pain from their GP each year (Arthritis Research Campaign, 2002).

Maniadakis and Gray estimated that with 37% of the population being affected by back pain each year, this represented a financial cost of £1,632 million to the exchequer, which rises to £12.8billion when costs from sickness absence and interventions are included (Maniadakis and Gray, 2000). There can be significant additional cost to the individual and their family where such symptoms are experienced in both the short and long-term. This can impact further, affecting ability to work, and in some cases, the advent of depression associated with long term pain and disability. The studies evaluated to calculate the economic burden for low back and musculoskeletal pain are now dated. In the intervening years since the estimates published by Maniadakis and Gray (2000) the economic climate in the UK has changed and inflation has affected the economy. The UK retail price index,

however, increased by 28.8% in the ten years to July 2008 (ONS, 2008) suggesting that current direct health care costs are likely to be substantially greater than the published figures. However, despite the fact that the UK has an increasingly aging population, there is no convincing evidence that age affects the prevalence of back pain (ONS, 2011; Airaksinen *et al.*, 2006) potentially adding to the cost estimates for back pain management. In 2013, 131 million days were lost due to general sickness in the population with 90.6 million due to neck, back, and muscle symptoms (Dunn *et al.,* 2013). This represents more days lost than for any other cause, and equivalent to 4.4 days per worker (ONS, 2014). In contrast, current estimates in Europe show the cost of back pain in Germany to be 0.9% of Gross Domestic Product (Wenig *et al.,* 2009), 1.7% GDP in Sweden, and 0.7% GDP in the Netherlands (Ekman *et al.*, 2005).

This significant costs to government and individuals highlights the need to focus on musculoskeletal symptoms in a sustainable manner other than through periodically funded RCTs. Current evidence suggests that there is a lack of consistency across PROMs use in the NHS and across professional organisations practising predominantly in the private sector (Unpublished data from qualitative interviews with osteopaths, chiropractors, and physiotherapists – see Chapter 3). To try to address this issue, a recent initiative in musculoskeletal care sponsored by Arthritis Care UK has focussed on the development of a generic musculoskeletal PROM (M-PROM) to be used pan-professionally (M-PROM briefing, 2013; M-PROM development, 2013). This measure, since renamed the Musculoskeletal Health Questionnaire (Msk-HQ), has undergone its initial development and has undergone reliability testing in a range of settings including a unit specialising in joint replacement surgery (hip, knee, and shoulder), a rheumatoid arthritis clinic, and a community primary care setting (Arthritis Care, 2015). It has recently been launched for wider use throughout the NHS, and in some manual therapy settings (Hill *et al.,* 2016). The routine use of PROMs in practice is being encouraged across manual therapy professions by a range of stakeholders including regulators, insurers, government, and by clinicians themselves to allow reflection on their own practice, and identification of areas of appropriate continuing professional development (CPD) (Stevens and Palfreyman, 2012; Tadić *et al.,* 2013).

The introduction of the Health and Social Care Act (2012) created an extensive reorganisation of the infrastructure of healthcare through the creation of NHS England (Figure 1.4). *Certain provisions of the Act extend to Northern Ireland, Scotland and Wales through territorial applications (Health and Social Care Act (2012) C7, Explanatory Notes).* It has, however, also produced opportunities for commissioning of services by manual therapists *e.g.* osteopaths. The delivery of such services, however, requires the provision of evidence either from traditional clinical trials or from PROMs.

1.2.4 International PROM initiatives

The use of PROMs has grown substantially in the UK within the past decade, but earlier international work is notable. In the USA, the National Institute for Health (NIH) allocated funding for the creation of the Patient Reported Outcome Measure Information Service (PROMIS) database. In their 2009 publication, Fries *et al*. reported on the NIH statement

"*In late 2004, a group of outcomes scientists from seven institutions and the National Institutes of Health (NIH) formed a cooperative group funded under the NIH Roadmap for Medical Research Initiative to re-engineer the clinical research enterprise. This initiative - the Patient-Reported Outcomes Measurement Information System (PROMIS) – aims to revolutionise the way patient-reported outcome tools are selected and employed in clinical research and practice evaluation. It will also establish a national resource for accurate and efficient measurement of patient-reported symptoms and other health outcomes in clinical practice*" (Fries *et al.,* 2009).

This is not the first initiative in the USA to try to develop a system for collection of PROM data. In his 1988 Shattuck lecture, Ellwood coined the term "outcomes management" (Ellwood, 1988). He stated that "*he envisaged a future in which patient management would be driven by the experience of how similar patients fared as a consequence of alternative treatments*" (Krumholtz, 2009). Wennberg *et al.,* 1993 describe another outcomes initiative where Patient Outcomes Research Teams (PORTs) were created for difference disease states *e.g.* diabetes and prostate disease.

Although value was gained for this initiative, when funding ended the project was largely discontinued.

Clancy and Eisenberg marked the entry of the term outcomes research into the scientific lexicon in 1998 with an article in the journal *Science* which stated that "*outcomes research - the study of the end results of health services that takes patients' experiences, preferences and values into account – is intended to provide scientific evidence relating to decisions by all who participate in healthcare*" (Krumholtz, 2009; Clancy and Eisenberg, 1998). These thoughts were supported by Porter, in a 1998 Congress appropriations hearing, who stated "*What we really want is to get at is not how many reports have been done, but how many people's lives are being bettered by what we have accomplished*" (Porter, 1998). Wressle *et al.* describe an initiative in Canada to create a "performance tool" to be used to collect ongoing data for patients with rheumatoid arthritis (Wressle *et al.,* 2003). The Canadian Occupational Performance Measure (COPM) was developed and tested for a day treatment programme; it was regarded as a successful national initiative to ensure patient compliance, involvement in the goal-setting process, and treatment planning and evaluation. It has the potential to be disseminated across a wider range of services in Canada.

Figure 1.4 Health and social care organisations in England (2016) based upon (http://www.nhs.uk/NHSEngland/thenhs/about/Documents/nhs-system-overview.pdf)

Abbreviations: HSCIC: Health and Social Care Information Centre, NIHR: National Institute for Health Research, NICE: National Institute of Health and Care Excellence, CQC: Care Quality Commission, MHRA: Medicines and Health Products Regulatory Agency, HRA: Health Research Authority, CCG: Clinical Commissioning Group, and GOsC: General Osteopathic Council.

1.2.5 PROMs in musculoskeletal care

The introduction of the Health and Social Care Bill (2012) afforded osteopaths and other healthcare professionals the opportunity to bid for different services in musculoskeletal care commissioned by local CCGs; such bids require a range of information and resources including the provision of PROM data. Consequently there is a clear need for a facility for osteopaths to routinely collect high quality outcome data to meet the challenge of this competitive commissioning process, but also to reflect on current standards of care and services. In response to this initiative (known as Any Qualified Provider) and other developments, a range of different resources are being developed to facilitate PROM data collection. In the chiropractic profession this has been achieved through the Care Response service (Care_response.org.uk, 2012); in physiotherapy via practice management software providers (CSP, 2012), and in primary care through a service called "myclinicaloutcomes.co.uk ". However, the NHS which represents the largest potential user of different types of PROM still utilises hard copy data collection. This labour- and resource-intensive process requires change but this will be hard to implement. Previous attempts at large scale Information Technology (IT) initiatives in the NHS have not been successful. Considerable thought, planning and resource allocation will be required to ensure PROM data collection occurs in a more successful manner (NHS connecting for Health, 2013; National Audit Office, 2014). In recognition of previous issues with large scale IT projects, the Msk-HQ (discussed earlier) will be released as a dual initiative employing both electronic and hard copy completion. The electronic version will be facilitated by ISIS Innovations in Oxford to facilitate this process (Personal communication with Jo Partington, PROM lead NHS England, 2014).

1.2.6 Different uses for PROMs

Although the term patient reported outcome measure, or PROM, is being used increasingly, it is important to be clear about its meaning and potential application to patient care. A PROM is essentially a form of questionnaire to measure a patient's health status. In osteopathic practice that measurement might include pain, disability, quality of life, fatigue or satisfaction: the key point is that this

measure is from the patient's perspective rather than from an osteopath or any other clinician.

PROMs can have a range of uses in clinical practice which include, for example:

- Measuring effectiveness of care;
- Fostering discussion with patients;
- Monitoring progress;
- Informed decision-making;
- Personalised care planning;
- Periodic review of treatment plans;
- Measuring overall population health;
- Allowing self-assessment (Greenhalgh *et al.,* 2009).

Historically outcomes of care were measured in terms of mortality and morbidity: the rationale for treatment interventions being based upon objective measures *e.g.* laboratory tests, radiographic evidence, and patient clinical evaluation (Fitzpatrick *et al.,* 1992). Fitzpatrick *et al*. identified seven major types of instrument for patient reported outcome measurement (PROM) including:

- disease-specific;
- site-specific;
- dimension-specific;
- generic;
- summary item;
- individualised;
- utility (Fitzpatrick *et al.,* 1998).

Patrick *et al.* writing in the Cochrane Review Group on PROMs identified some key points:

- PROMs are reports coming directly from patients about how they feel or function in relation to a health condition and its therapy without interpretation by healthcare professionals or anyone else.

- PROMs can relate to symptoms, signs, functional status, perceptions, or other aspects such as convenience and tolerability.
- Items reflecting the concepts included in a PROM questionnaire are elicited from the target population.
- Patient involvement in questionnaire generation is essential for content validity.
- PROMs are not only important when more objective measures of disease outcome are unavailable but also to represent what is most important to patients about a condition and its treatment.
- PROMs can be continuous (*or more correctly quasi-continuous*) or categorical. Techniques are available to pool both kinds of measures (Patrick *et al.,* 2008).

In contrast, PREMs are less well developed in healthcare. In some settings the PREM encompasses a range of issues relating to a clinical setting including elements of the site content in addition to service delivery information. Coulter *et al.,* 2009 stated that a number of studies have reported improvements following systematic gathering of patient feedback by hospitals (Draper *et al.,* 2001; Hildenhovi *et al.,* 2002; Crawford *et al.,* 2002; Gillies *et al.,* 2003; Reiber *et al.,* 2004; Sweeney *et al.,* 2005; Davies and Cleary 2004; Richards and Coulter 2007; Davies *et al.,* 2008; Bate and Robert 2006; Forbat *et al.,* 2009), but in general this has not been given high priority in NHS organisations. Despite the requirements of clinical governance and the emphasis on patient-focussed care, it is still rare for patients to be asked routinely to comment on the quality of their care, and for the most part any quality improvements resulting from feedback have been small. Failure to listen to patients' and relatives' criticisms has been implicated in investigations as a key factor in failing hospitals (Department of Health 2001; Colin-Thom é, 2009). A recent study suggested that achieving and sustaining more substantial change is likely to require organisational strategies, engaged leadership, cultural change, and regular measurement and performance feedback (Davies *et al.,* 2008). While in osteopathic practices such changes may be easier to instigate, the lack of resources in terms of available time and staff may hinder their implementation.

Outcomes can include a range of different metrics and domains. The key measurements recommended by NHS England include:

- Patient Reported Outcome Measurement (PROM);
- Patient Reported Experience Measurement (PREM);
- Patient satisfaction;
- Global impression of change.

These measurements will be considered in greater detail in the Chapters 4 and 5. More recently, PROMs have evolved to include more focus on individual patients and their goals, and Patient Centred Outcome Measures (PCOMs) have been created. The location of outcomes within the four domains of quality metrics can be seen in Figure 1.5 (Blount *et al.,* 2012).

Traditionally clinicians have tried to determine a patient's response to treatment by measuring a range of different items which may be associated with issues raised by the patient during the initial consultation, or based upon clinical measures assessed during physical examination. They are sometimes viewed as surrogates to infer functional ability (Dettori, 2007). These types of measure have been described in the literature as clinician based outcomes (CBOs).

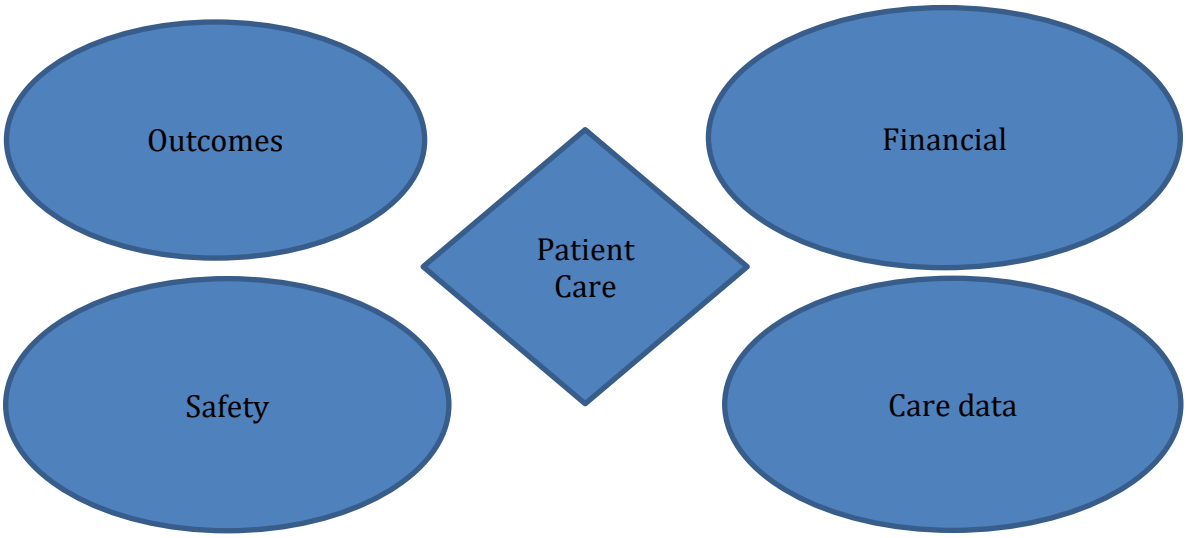


Figure 1.5. Four Domains of Healthcare Metrics.
Adapted from Blount *et al.,* 2012.

In addition to physical measures, CBOs can also refer to different tests that are used to support the assessment of a patient's health. Examples include gait measurement, responses to provocative tests, and muscle strength. Quantitative measurement of the types of CBOs listed have traditionally been considered as more "objective", however this stance is challenged by many published papers which address inter-rater and intra-rater reliability and find them notably lacking in agreement (Moran and Gibbons, 2001; Edwards *et al.,* 2002).

While the clinician has traditionally been the arbiter of decision-making about patients' care, the political landscape has changed increasingly, and within this the role of the patient in healthcare has changed too.

1.2.7 The changing role of the patient

The impetus for developing patient-focussed studies originally began in the 1970s; principally in the USA. This occurred due to government support and an increased interest in the quality of medical care (Ware and Snyder, 1973). The Griffiths report (1983) encouraged the role of the consumer as a legitimate judge of quality and called for measurement of levels of satisfaction through patient surveys (Magni *et al.,* 1993; Croft *et al.,* 1997). There has been an increased shift in consumerism and a consumer-orientated culture in healthcare in the interests of maintaining a competitive edge (McIver, 1991a; McIver, 1991b); the term consumer has appeared increasingly in UK patient satisfaction literature (Hopkins, 1990; Williams and Calnan, 1991a; Cox *et al.,* 1993; Hudak, 2003). Thompson described prevailing models of patient involvement in care and the consequent shift in the balance of power:

- paternalism (involvement limited to receiving information or giving consent);
- shared decision-making (options are shared between patient and practitioner);
- practitioner-as-agent (practitioner holds technical expertise, but patient preferences are incorporated into decision-making);
- informed decision-making (technical expertise transferred to patient who makes the final decision) (Thompson, 2007).

Such notions have proved challenging to established NHS culture, with "recognition of the patient as a stakeholder rather than a grateful recipient in the provision of healthcare" requiring adjustment to both the process and environment of service delivery (Arnold, 2004). There is, however, a contrast between NHS and private sector healthcare; "Patient power should be no more problematic within an NHS system than it is in a system of health provision in which the patient is a paying client" (Arnold, 2004). In a competitive market place it is imperative for private sector practitioners to retain existing patients/clients and recruit new ones. Since word-of-mouth recommendation is an important source of self-referral for fee-paying clients (Potter *et al.,* 2003), failure to recognise their power as stakeholders and the potential impact on their ideal, predicted or normative expectations, could have implications for business success (Leach *et al.,* 2013).

Increasingly patient-focussed care has been advocated by healthcare providers both in the UK and internationally, but it has also become an implicit requirement of clinical governance. The introduction of clinical governance into the healthcare arena has affected not only NHS practitioners but those in complementary health care professions such as osteopathy (Scally and Donaldson, 1998; Wilkinson *et al.,* 2004). Clinical governance has modified the focus from quality assurance to encompass standards on record keeping, monitoring outcomes, clinical audit, patient satisfaction measures, patient safety, and the implementation of evidence (Anonymous, 1989). These demands reflect some of the requirements outlined in the recent "Osteopathic Practise Standards" issued by the General Osteopathic Council (GOsC, 2012). Patient satisfaction (PS) constitutes one aspect of clinical governance and studies examining PS have been undertaken in a range of different settings. These settings have included osteopathic clinics attached to osteopathic educational institutions (OEIs), and osteopathic services provided on GP premises. One study within an OEI was a descriptive and exploratory investigation of patient satisfaction and perceptions of treatment. The majority of patients expressed satisfaction with treatment, the explanations they received and their perceived health outcomes (Strutt *et al.,* 2008). Chronic low back patients reported their satisfaction with treatment they received for back pain from GPs and osteopaths

practising within the same surgery. Although levels of satisfaction were high for all treatments, patients reported significantly higher scores for satisfaction with the osteopathic treatment (Pincus *et al.,* 2000).

1.2.8 Evidence for osteopathic practice

Osteopathy, Physiotherapy, and Chiropractic are three professional groups largely concerned with the therapeutic management of musculoskeletal presentations. In the literature the three professions are commonly referred to collectively as "manual therapy" professions: although they hold different philosophical viewpoints they share more similarities than differences in their approaches to patient management (Carnes and Fawkes, 2013). There is increasing interest in the provision of osteopathy from the public at large, from the NHS and from government (Thomas *et al.,* 2003; House of Lords, 2000). This type of treatment is currently used by some 13% of the population in the UK (Thomas *et al*., 2001).

Osteopathic care contains over 100 different techniques or procedures (Lesho, 1999; DiGiovanna *et al.,* 1991; Greenman, 1989; Still, 1992; Owens, 1963). The most commonly used structural approaches are broadly grouped into seven major types:

- High velocity low amplitude (also called thrust or manipulation techniques). This involves a quick movement within a joints normal range of movement and does not exceed the anatomic barrier of the joint. Movement can be targeted to specific spinal segments and, with appropriate positioning of the patient, requires very little force. The goal of the technique is to restore joint play (Heilig, 1986; Chila *et al.,* 1990). The technique is frequently characterised by a clicking sound whose source has been investigated by a number of researchers (Evans, 2002; Evans and Breen, 2006). This technique most closely resembles chiropractic manipulation and is subject to most contraindications;

- Soft tissue/massage techniques (Furlan *et al.,* 2009);

- Articulation involving gentle repetitive movement of a joint to try and

increase the range of movement (Heilig, 1961);

- Muscle energy technique. This involves repeated isometric contractions with passive joint movement to increase joint mobilisation and lengthen contracted muscles (DiGiovanna *et al.,* 1991; Greenman, 1989);

- Counterstrain techniques. This involves the symptomatic joint being placed in a position of least discomfort while at the same time monitoring the degree of tenderness at a nearby tender point until the tenderness reduces (DiGiovanna *et al.,* 1991; Greenman, 1989; Jones, 1981). The only contraindication is patient unwillingness or inability to cooperate;

- Myofascial release techniques. These techniques are similar to deep massage techniques and are designed to stretch muscle and reduce tension (Greenman, 1989);

- Lymphatic pump techniques. These techniques attempt to mechanically assist lymphatic drainage. There are a small number of contraindications to this technique (Degenhardt and Kuchera, 1996).

The wide range of techniques ensures that care of the patient is tailored to their general health and wellbeing, their age, presenting symptoms and any comorbidities. A wide range of symptoms are treated in clinical practice; low back pain is the most common but pain to the cervical spine, shoulder joint, and knee joints are also very commonly presented (Fawkes *et al.,* 2014b).

Access to osteopathic treatment is through a variety of locations: private practices, NHS hospital outpatient departments, General Practices (GPs) and clinics attached to osteopathic education institutions (Thomas *et al.,* 2001; Ong, 2004; Wye *et al.,* 2009; Langworthy *et al.,* 2000). However, the vast majority of patients access treatment through private practices. Traditionally, musculoskeletal disorders, particularly low back pain, have been the most common reasons for a patient to visit an osteopath. The limited survey work that has been done suggests that back pain accounts for approximately fifty percent of an osteopath's workload and musculoskeletal-type presentations make up the majority of the rest of the

caseload (McIlwraith, 2003). This lack of basic survey work about the profession and its inability to describe itself led to a series of innovations. Data were collected in a number of locations and in a variety of different ways including retrospective case note review (Burton, 1981), in teaching clinics (Hinkley and Drysdale, 1995), single practices (McIlwraith, 2003; Pringle and Tyreman, 1993), or single day surveys with poor response rates (GOsC 1998; GOsC, 2001). A new initiative began in 2007 with the development of a standardised data collection (SDC) tool for the osteopathic profession to collect data on patient profiles, their route to treatment, the interventions delivered, and outcomes of care. Data collection to pilot the SDC occurred over a three month period in 2009 (Fawkes *et al.,* 2014a; Fawkes *et al.,* 2014b). While these data collection initiatives have yielded considerable amounts of useful information, they have failed to deliver compelling and robust, unbiased data concerning outcomes of osteopathic care. The necessity for a system to allow collection of ongoing patient completed data was highlighted by all initiatives. The issue of various forms of bias being inherent in outcome data collected directly by clinicians was one of the key drivers for a system to allow patient completion away from a practice location, and for not allowing osteopaths to see individual patient data, only summary data (Pannucci and Wilkins, 2010).

Alternative sources of data exist in terms of clinical trials and subsequent reviews and meta-analyses but the quantity of this material is limited due to access to funding to initiate costly clinical trials. Despite its limited evidence base, osteopathy increasingly features in clinical recommendations, notably for back pain. The Clinical Standards Advisory Group (CSAG) produced clinical guidelines for the management of acute low back pain, in 1994, which produced guidance on diagnostic triage, and principal recommendations for treatment based on evidence in this area (CSAG, 1994; RCGP, 1999). Manipulation was recommended “within the first six weeks of the occurrence of symptoms for patients who need additional help with pain relief or who are failing to return to normal activities”. The European low back pain guidelines (ELBPG) project group, COST B13, examined both acute and chronic back pain and made recommendations accordingly. Their acute low back pain guidelines suggest "consideration of referral for spinal manipulation for patients with acute low back pain who are failing to return to

normal activities" (ELBPWG, 2004a). The guidelines for chronic low back pain recommend that "short courses of manipulation/mobilisation can also be considered for chronic low back pain patients (ELBPWG, 2004b).

This work was followed by the Musculoskeletal Services Framework which provides advice concerning the use of osteopathic care/spinal manipulation (Department of Health, 2006). NICE has reviewed the evidence looking at the management of subacute non-specific low back pain; this looks specifically at back pain that has lasted longer than six-weeks but not more than thirteen-months (Savigny *et al.,* 2009). The consultative process began in 2008 and guidelines were produced in May 2009. The guidelines have produced information concerning a variety of different treatments and approaches for patients with non-specific low back pain including up to nine sessions of manual therapy treatment which includes osteopathy. These guidelines have been poorly implemented and are currently being updated (publication due November, 2016) (BOA, 2009; NICE, 2016). A wide range of osteopathic trials have been published of varying quality; some of these, in combination with studies from the medical, physiotherapy and chiropractic professions, were examined by NICE during the guideline development process.

When reviewing clinical trials it should be remembered that low back pain is the symptom for which the highest numbers of patients consult osteopaths (GOsC, 2001). Commentators have recorded the view that for acute uncomplicated low back pain “osteopathy and chiropractic were rated as effective by most experts” (Ernst and Pittler, 1999). Osteopathic trials have looked at the management of patients with acute low back pain (Gurry *et al.,* 2004), acute and chronic (Hoehler *et al.,* 1981; Gibson *et al.,* 1985; Andersson *et al.,* 1999; Williams *et al.,* 2003).

The management of patients with chronic low back pain was the focus of the 2004 Medical Research Council-funded United Kingdom Back Pain, Exercise and Manipulation (UK BEAM) randomised trial (UK BEAM Trial Team, 2004a). This looked at how a package of care involving one or a combination of treatment approaches could improve low back pain in patients. The study’s authors concluded that the combination of spinal manipulation and exercise was more

beneficial than when the treatments were used in isolation, and when compared to "best care" offered through general practice. An economic evaluation was made for this study and this concluded that adding spinal manipulation to "best care" was a cost effective way to manage back pain in general practice (UK BEAM Trial Team, 2004b). Further analysis of the BEAM trial data has recently been undertaken looking specifically at the number needed to treat (NNT) (Froud *et al.,* 2009). The NNT is an indication of the number of patients who will need to receive treatment with a specific intervention for one to improve. This work found that, in contrast to the small mean differences originally reported in the BEAM trial data, NNTs were small. Froud *et al*. identified that at three months, NNT estimates ranged from 5.1 (95% CI 3.4 to 10.7) to 9.0 (5.0 to 45.5) for exercise, 5.0 (3.4 to 9.8) to 5.4 (3.8 to 9.9) for manipulation, and 3.3 (2.5 to 4.9) to 4.8 (3.5 to 7.8) for manipulation followed by exercise. The low number of patients who would need to receive spinal manipulation to gain a beneficial change could be attractive to clinicians, patients, and purchasers. Further analysis of the BEAM trial data has attempted to identify characteristics of randomised controlled trial participants which predict greater benefits from physical treatments for low back pain: in turn this would allow more appropriate selection of patients for different treatments (Underwood *et al.,* 2007). The analysis of this data found that baseline participant characteristics did not predict response to the UK BEAM treatment packages, and in particular, this analysis suggested that the distinction between sub-acute and chronic low back pain may not be useful when considering treatment choices.

In the USA, work was also undertaken by Licciardone *et al*. investigating osteopathic treatment of patients with chronic low back pain (Licciardone *et al.,* 2003). Further work looking specifically at biomechanical measures and somatic dysfunction has recently been published by the same research team (Licciardone *et al.,* 2008; Licciardone *et al.,* 2014a; Licciardone *et al.,* 2014b). Licciardone *et al*. report that at baseline, prevalence rates of non-neutral lumbar dysfunction, pubic shear, innominate shear, restricted sacral nutation, and psoas syndrome were determined in 230 patients. Each patient received five OMT (osteopathic manipulative treatment) sessions which were delivered at weeks 0, 1, 2, 4, and 6. The prevalence of each of the biomechanical dysfunctions cited above was again

measured at week 8 before the final OMT session took place. Significant improvements in each biomechanical dysfunction were observed with OMT.

Other research has been conducted looking at specific patient populations. This has included patients with disc injury and their treatment by either spinal manipulation or chemonucleolysis (Burton *et al.,* 2000), patients during pregnancy (Sabino and Grauer, 2008; Licciardone *et al.,* 2009), and patients with psychological disorders (Pincus *et al.,* 2002; Williams *et al.,* 2007). Sabino and Grauer, 2008, and Licciardione *et al.,* 2009 undertook studies showing back pain during pregnancy decreased with usual obstetric care and osteopathic manipulative treatment, remained unchanged with usual obstetric care and sham ultrasound treatment, and increased in the usual obstetric care only group, although no between-group difference achieved statistical significance. The researchers concluded that osteopathic manipulative treatment slows or halts the deterioration of back-specific functioning during the third trimester of pregnancy. In their systematic review and meta-analysis of studies involving spinal manipulation where a psychological outcome was measured, Williams *et al*. identified twelve trials in total: six trials with a verbal intervention comparator were combined in a meta-analysis (Williams *et al.,* 2007). They found a mean benefit from spinal manipulation equivalent to 0.34 of the population standard deviation (SD) at 1-5 months; 0.27 of the SD at 6-12 months. Eight trials with a physical treatment comparator were combined in a meta-analysis and found a mean benefit of 0.13 of the SD in favour of manipulation at 1-5 months; 0.11 of the SD at 6-12 months. The researchers concluded that there was some evidence that spinal manipulation improved psychological outcomes compared with verbal interventions. The work by Burton *et al*. in 2000 compared the use of spinal manipulation with chemonucleolysis in the management of patients with disc herniation. Osteopathic manipulation produced a 12-month outcome that was equivalent to chemonucleolysis, and at the time it was suggested as an option for the treatment of symptomatic lumbar disc herniation where indications for surgery were absent. However, chemonuleolysis is no longer offered as an intervention.

As the body of osteopathic trial literature has grown, this has allowed the conduct of systematic reviews, meta-analyses, a comparative effectiveness meta-analysis, and guideline development by a range of different authors. It is here that the disagreements concerning the effectiveness of manual therapy interventions are most clearly highlighted with reviews supporting the use of manual therapies for spinal pain (Assendelft *et al.,* 2003; Liccairdone *et al.,* 2005; Savigny *et al.*, 2009; Brønfort *et al.,* 2010; Clar *et al.,* 2014), and those disputing their effectiveness (Posadzki and Ernst, 2011a; Posadzki and Ernst, 2011b; Menke, 2014). The effect of environment and access was investigated by Chown *et al.,* 2008. Outcome of care was measured when comparing one-to-one osteopathic sessions with a group exercise class. Attendance at the one-to-one sessions was 80%, and mean change in the score of one of the outcome measures used (the Oswestry Disability Index) was greater by 0.84 for individual treatment sessions compared to group exercise classes.

While clinical trials remain an important part of the evidence base for any healthcare profession, there remains a need for ongoing patient data to be collected to describe professions more robustly and inform the need for future clinical trials. Ongoing data collection in the form of PROMs is an area of practice termed clinimetrics. This area of practice has its own terminology and requirements which make it valuable to patient care.

Measurement in healthcare

The science of clinimetrics has helped to increase the patient's voice in the measurement of their health. The traditional (bio)medical model of health was developed based on the work of scientists in the 19th Century (Pasteur, 1858; Cohn, 1856; Koch, 1882; Koch, 1884). During the latter stages of the 19th and 20th centuries, the physical evidence for some diseases was identified leading to the development of effective treatments. It is from this basis that the medical model emerged where the concepts of disease and injury have a central role. More recently, the term "medical model" has been used and attributed to Laing who described it as the "set of procedures in which doctors are trained" (Laing, 1971). The term "medical model" has been defined in greater detail as "*a scientific process*

involving observation, description, and differentiation, which moves from recognising and treating symptoms to identifying disease aetiologies and developing specific treatments" (Clare, 1980; Shah and Mountain, 2007). While it is regarded as the predominant western approach to illness by many commentators, it has been challenged by other philosophical and cultural approaches. Most notable among the challengers has been Engel. In his seminal paper "The need for a new Medical Model: A Challenge for Biomedicine", he stated

"Over the past 50 years, medical education has grown increasingly proficient in conveying to physicians sophisticated scientific knowledge and technical skills about the body and its aberrations. Yet at the same time it has failed to give corresponding attention to the scientific understanding of human behavior and the psychological and social aspects of illness and patient care" (Engel, 1978).

Engel proposed that the social and psychological world of an individual encompassing thoughts, behaviours, and feelings is as worthy of study in relation to health and illness as the biological factors underpinning disease development. While the biopsychosocial model, as with many new models of health, could be challenged for lack of underpinning evidence, it could be rebutted by the evidence that among the 30% of cancers may be associated with tobacco use (lung) or diet (digestive tract) (Doll and Peto, 1981), psychological factors are recognised as markers for pursuing healthy lifestyles (McInerny, 2015). At the other end of the coping spectrum is the reaction to severe stress which provides a typical and very common example of biopsychosocial disorder (Lazarus and Folkman, 1984; McInerny, 2015). The original version of the Biopsychosocial model included three areas of focus (biology, psychology, and sociology). This has been expanded more recently to encompass the role of cultural and spiritual factors in the health and wellbeing of an individual (Meinert and Yuen, 2012) as shown in Figure 1.6.

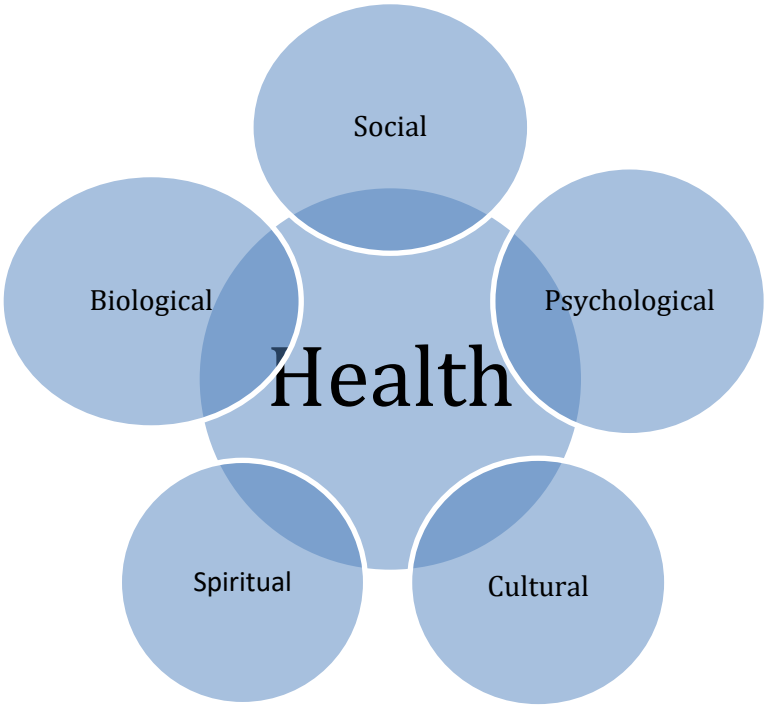


Figure 1.6. The biopsychosocial model of health and wellbeing.

The areas of focus within the Biopsychosocial model not only have relevance to clinical practice, but also how we measure the effect of clinical practice upon the patient, their symptoms, and their experience of care. This will be discussed in greater depth in Chapter 4 which addresses the content of Patient Reported Outcome Measures (PROMs).

1.2.9 Cinimetrics in healthcare

Clinimetrics is the science of measurement that underpins many Patient Reported Outcome Measures (Feinstein, 1967). In common with any discipline, clinimetrics has its own unique terminology. One issue that has arisen historically is that different groups involved in this area of study use different terms to describe the same thing. Work has been ongoing between different academics to try and address this situation and reduce confusion for clinicians and other professionals who may be using PROMs. The area of terminology has not reached full consensus within the scientific community specialising in clinimetrics/psychometrics (Mokkink *et al.,* 2006). The definitions described below are based on the terminology used by the **CO**nsensus-based **S**tandards for the selection of health **M**easurement **In**struments (COSMIN) group (http://www.cosmin.nl/) in the Netherlands. The taxonomy described by the COSMIN group is summarised

visually at http://www.cosmin.nl/images/upload/files/COSMIN%20taxonomy.pdf.

Feinstein defined clinimetrics as "The practice of assessing or describing symptoms, signs, and laboratory findings by means of scales, indices, and other quantitative instruments" (Feinstein, 1982). Healthcare involves several different forms of metrics and there is some agreement that the term clinimetric should be reserved for multidimensional health measurement scales and indices. A summary guide to the different definitions that will be adopted throughout this thesis are provided overleaf. They are summarised diagrammatically in Figure 1.7. A full explanation of clinimetric terminology will be included in Chapter 4 of this thesis.

While different groups are exploring different terminology in clinimetrics, other groups have focussed their attention on gaining consensus among healthcare professionals on the core outcome measures to be used within different specialist areas in health. When addressing medicine and healthcare generally, the COMET-initiative has attempted to identify a core set of PROMs: in musculoskeletal care different consensus groups including IMMPACT, MMICS (Pincus *et al.,* 2008), OMERACT, and ICHOM have attempted to agree on core measures to be used.

1.2.10. Methods for health evaluation

Questionnaires measuring health status or the outcome of care have been used in clinical research and epidemiological studies for many years. The questionnaire was invented originally by the Statistical Society of London in 1838 (Gault, 1907), and is now delivered in a range of formats using paper, telephone, or technological devices (Hox and de Leeux, 1994).

There are pros and cons to all approaches, and areas for consideration when adopting each approach. The literature on this area is growing although comparisons of paper versus electronic delivery of questionnaires are limited despite this becoming an increasingly important area of research as data collection generally and PROM data specifically move to electronic formats.

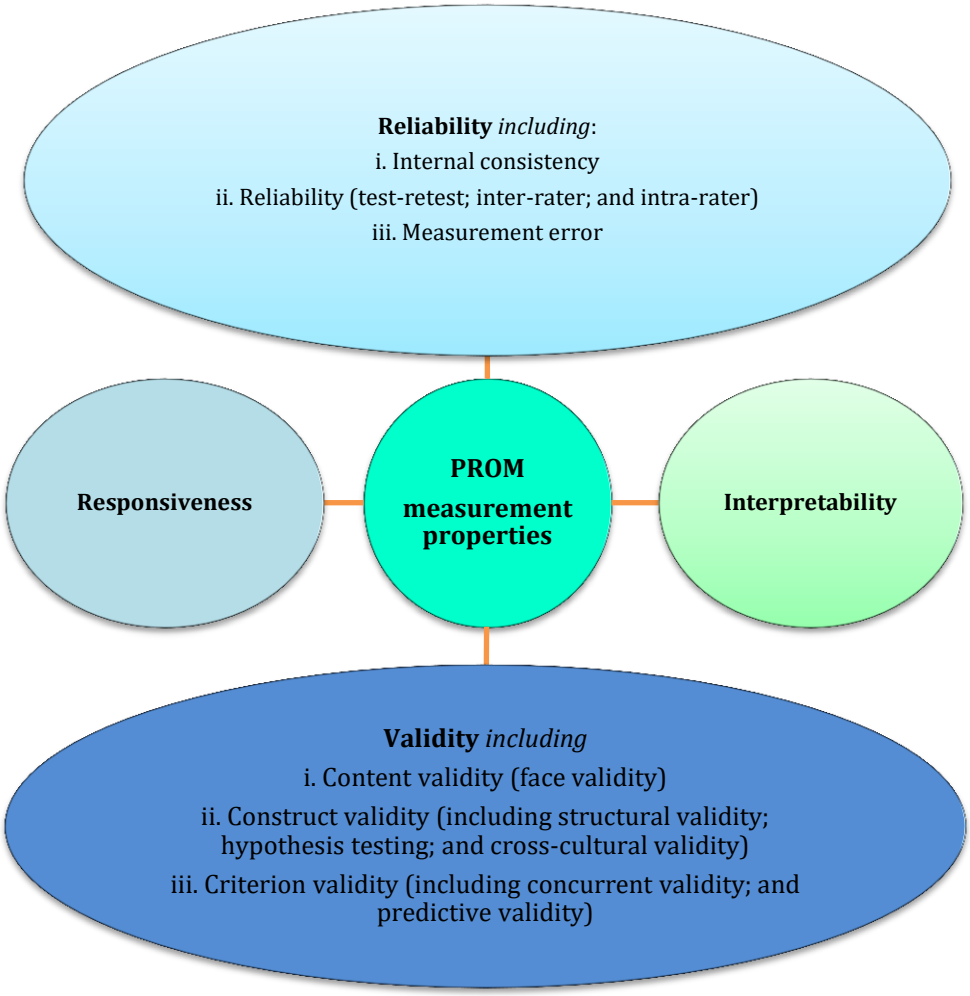


Figure 1.7 A summary of clinimetric properties of a Patient Reported Outcome Measure (PROM) based on COSMIN taxonomy (http://www.cosmin.nl/).

When available resources are unlimited, the consideration whether to use paper, electronic, or telephone data collection is less pressured. However, health budgets are increasingly constrained as health spending for 2014-2015 reached £113.3 billion with a deficit for the NHS of £471million (NHS England, 2015a; NHS Confederation, 2016). Even with a budget of £116.4 billion for 2015-2016, economies in clinical areas will be required increasingly and this will include evaluation of care (NHS England, 2015b). While previous studies have indicated the popularity of paper-based questionnaires compared to electronic, and their associated higher response rates, much of the literature in this area is growing quite dated (Couper, 2005; Dillman, *et al.,* 2007).

However, while the response rates for questionnaires are important, of equal consideration is the response representativeness for the population being investigated (Cook *et al.,* 2000). Factors investigated to increase improved response rates concern colour of questionnaire, provision of stamped envelopes, postal reminders, and incentives for questionnaire completion. Some of these factors are relevant to electronic questionnaires also, but electronic surveys avoid the administration errors and costs associated with postal surveys (Dommeyer *et al.,* 2004; Nulty, 2008). Schaefer and Dillman identified that email pre-notice was more effective in increasing response rate than mail pre-notice, but multiple notices failed to have greater effect (Schaefer and Dillman, 1998; Kaplowitz *et al.,* 2004). When making comparisons between the two modes of questionnaire completion, Dommeyer identified a 43% response rate for electronic questionnaires, and 75% for paper questionnaires (Dommeyer *et al.,* 2004). In his more recent evaluation, Nulty assessed comparative studies from 1999 to 2006, and identified an overall response rate of 56% across all studies for paper completion, and 33% for electronic completion (Nulty, 2008). This difference of 23% is comparable to the meta-analysis by Shih and Fan who also identified a 20% difference in favour of paper questionnaires (Shih and Fan, 2008). However, the timing of the studies should add caution to the interpretation of such findings. In the intervening period since these papers were published access to the Internet has increased in households, and in other forms, *e.g.* public libraries (Hohwü *et al.,* 2013; Gartner, 2016). In other studies where preference between paper and electronic questionnaire has been investigated specifically, growing numbers of participants preferred electronic questionnaires across all age groups *e.g.* 69% of adults, 77% of children, and 73% of care givers (Bushnell *et al.,* 2003).

1.2.11 The changing role of the clinician

I described in Section 1.2.1 how the osteopathic profession has developed since osteopathy's introduction into the UK. While professional regulation has raised the profile of the profession this also brings responsibilities in the form of practise standards (GOsC, 2012). The document *Osteopathic Practice Standards*, (2012) makes clear some requirements concerning gathering patient feedback, and assessing quality of practice through activities including data collection and clinical

audit. More recently, the focus has been upon PROMs in osteopathic practice. The profession as a whole represents a spectrum of osteopaths who graduated with a D.O. to those graduating with an MSc award. The capabilities within this spectrum concerning data collection (including PROM data), clinical audit, and research activities are varied (KPMG, 2011). When focussing particularly upon PROMs, there is no available research concerning osteopaths' views on the merits of PROMs or lack of them in osteopathic practice. While there is a growing body of literature from other professional groups concerning views and experiences about PROMs, there is nothing which has addressed the views of clinicians in private healthcare practice.

Notwithstanding this lack of osteopathic-focussed literature, there is a considerable range of views from other professional groups. Qualitative and quantitative approaches have been used to identify clinicians' views on PROMs. In their systematic review of qualitative studies concerning PROM barriers and facilitators, Boyce *et al*. identified four main themes including practical aspects (collecting and incorporating data), attitudinal (the value of the actual data collected by PROMs and its application to clinical practice and patient management), methodological (knowing what to do with the data to interpret it), and impact (using the data to make changes in patient management) (Boyce *et al.,* 2015). Although research has explored the views of other professional groups *e.g.* renal specialists (Breckenridge *et al.,* 2015) occupational therapists, physiotherapists, speech and language therapists (Duncan and Murray, 2012) athletic trainers (Valier *et al.,* 2014), unspecified experts, clinicians and managers (Van der Weees *et al.,* 2014), and physicians, nurses, and therapists (Boyce *et al.,* 2015) there is considerable overlap in factors that continue to emerge despite the different specialist areas.

Clear benefits have been identified in some studies many of which centre around enhanced communication with the patient whether that involves engagement in the therapeutic process, improving direction of patient care, or motivating and encouraging patients (Velicova *et al.,* 2004; Hatfield *et al.,* 2007; Jette *et al.,* 2009, Swinkels *et al.,* 2011, Valier *et al.,* 2014). Enhanced communication with other

healthcare professionals was also cited by other studies (Valderas *et al.,* 2008a; Kotronoulas *et al.,* 2014; Valier *et al.,* 2014). In addition, technical aspects of care were highlighted including being more thorough in examinations, supporting clinical reasoning, and clearer identification of treatment goals and interventions which could support these, and the opportunity for national benchmarking (Pisoni *et al.,* 2008; Swinkels *et al.,* 2011; Valier *et al.,* 2014; Breckenridge *et al.,* 2015).

1.2.12 PROMS and their effect on patient care

In their review of patient reported information (PRI) and its effect on clinical practice, Schlesinger et al. identified four different forms of patient information (Schlesinger et al., 2016). These included:

- Patient reported outcome measurements;
- Surveys of patient experience;
- Narrative accounts of patients' encounters with staff;
- Complaints and/or grievances.

This type of information is being collected increasingly frequently in a range of settings. In their earlier systematic review, Valderas et al. identified that the most common therapeutic areas where outcome measurement have been studied for their effect on clinical practice were mental health and oncology (Valderas et al., 2007). Their investigation of 19 RCTs in primary care identified that 65% of outcomes measured actual processes of care while only 47% measured outcome(s) of interventions. Among those studies that identified PROMs' use having a positive effect on outcome, the detection of physical or psychological problems that might otherwise be overlooked was highlighted (Lohr, 1992; Greenhalgh and Meadows, 1999; McHorney, 1999; Espallargues *et al.,* 2000). The monitoring of disease progression using standardised measures provided information on the impact of treatment on patients; this was supported also by improvement in clinical notation in notes (Rubinstein *et al.,* 1989). This was echoed in the word by Luckett *et al.* where they found that the process of care was improved for patients attending oncology clinics and mental health clinics respectively, but were less convinced that PROMs had an effect on outcome measured using satisfaction and quality of life measures (Marshall *et al.,* 2006; Luckett *et al.,* 2009). This view was echoed in

later work by Howell *et al.* while Chen *et al.* in their review of oncology services recommended increased use of PROMs to identify and support better patient-centred care (Chen *et al.,* 2013; Howell *et al.,* 2015).

In other studies, the improvement of patient-clinician communication and shared decision-making has been cited (Greenhalgh *et al.,* 2005). These effects must be qualified by identifying and observing shared priorities among patients and clinicians in planning care (Rothwell *et al.,* 1997). Identifying priority areas for patients and building these into care planning demonstrated also an enhanced effect in compliance with treatment and/or following advice (Stimson, 1974; Valderas *et al.,* 2008).

The potential adverse effects of PROMs' use have been studied. Valderas *et al.* suggest that for some patients the identification of physical or psychological symptoms of which patients had been unaware may cause greater concern having a potentially detrimental effect on wellbeing (Valderas *et al.,* 2008). This view has been supported more recently by Wolpert particularly in mental health settings (Wolpert, 2014). Wolpert's views were based on a range of issues including the use of questionnaires lacking relevance to the particular setting, lack of integration within the consultation, staff concerns about targets for completion of PROMs, lack of identification concerning how the PROMs add value to a service or intervention, concerns among patients and clinicians alike about how the use of PROMs may limit future service provision and potentially increasing the burden on clinicians by detecting problems which might otherwise not be detected (Ford *et al.,* 2009; Tavabie, 2009; de Jong *et al.,* 2012; Moran *et al.,* 2012; Wolpert, 2014).

While much of the literature understandably focusses on the effects of PROMs' use on patient care, their effect on clinicians has been investigated also. Clinicians have identified the importance of receiving feedback from PROMs in a manner which is both timely and comprehensible (Meadows *et al.,* 1998; Morris *et al.,* 1998; Boyce *et al.,* 2014; Kendrick *et al.,* 2016). This, in turn, allows patient care to be modified as necessary when PROMs are used as a monitoring tool, specifically for care planning, or as a decision aid. Although Greenhalgh and Meadows

identified that PROMs could increase detection of issues and facilitate communication, they found little evidence that this did have an effect on management or outcome (Greenhalgh and Meadows, 1999). The latter finding was supported by Gilbody *et al*. who regarded PROM use as insufficient to effect an improvement in outcome to mandate their use in a more widespread manner (Gilbody *et al.,* 2001). The findings of Espallargues *et al*. were more equivocal: their review stressed the heterogenous nature of the research in terms of settings, patient populations, clinical disorders, interventions, research design, and frequency of PROMs' use to draw any definitive conclusions (Espallargues *et al*., 2001). A lack of consistency with which measures were selected, and delivered/administered was noted also (Kazis *et al.,* 1990; Magruder-Habib *et al*., 1990; Wagner *et al.,* 1997).

Although the literature to date indicates areas where the use of PROMs can support clinical practice, and have a beneficial impact on patient care, there are some cautionary words from other investigators and clinicians. Sanata and Feeney recognising the challenge of introducing PROMs into patient care created a conceptual framework to develop implementation strategies especially for patients with ling terms conditions (Santana and Feeny, 2014). In the most recent work in this area, a realist synthesis undertaken by Greenhalgh et al. made a number of key recommendations for consideration concerning PROMs use. These recommendations included:

- The importance of giving feedback of aggregate PROMs and performance data to improve patient care;
- The importance of giving feedback of individual PROMs data when monitoring patients to improve patient care;
- The importance of examining how PROMs data can be implemented into clinical processes;
- Consideration of how PROMs feedback may challenge existing evidence for patient management;
- Encouragement of PROMs use as a means of supporting patients to raise or share their concerns;

- Examination by providers about how PROMs data can modify services to improve patient care (Greenhalgh *et al.,* 2017).

Earlier work has identified the heterogeneous nature of PROMs research, and further consideration needs to be given to the methodological approaches used in future to measure the impact of PROMs on clinical practice and patient care.

Barriers for the use of PROMs centre around a range of key themes. Murray and Duncan identified four including knowledge, education, and perceived value of PROMs; support/priority for PROM use; practical considerations; and patient considerations (Murray and Duncan, 2012). There is considerable overlap from other studies who have also cited time consumed for patients and clinicians (McAuley *et al.,* 2014; Valier *et al.,* 2014), confusion for patients and clinicians to use (Hatfield and Ogles, 2007; Valier *et al.,* 2014), lack of training for clinicians (Duncan and Murray, 2012; Valier *et al.,* 2014), being subjective to be useful (McAuley *et al.,* 2014; Valier *et al.,* 2014), increasing anxiety for patients (Valier *et al.,* 2014), requiring too high a level for reading and language ability (Valier *et al.,* 2014), lack of confidence in data analysis and interpretation (van der Wees *et al.,* 2014), and lack of cultural and ethnic considerations/sensitivities in the PROMs (Valier *et al.,* 2014). Despite the barriers and facilitators identified, Murray and Duncan noted that the lack of perceived value by clinicians was bi-directional when it came to implementation: there was a decreased likelihood of use where value was perceived to be high, but greater perceived value appeared to increase uptake (Russek *et al.,* 1997; Copeland *et al.,* 2008; Skeat *et al.,* 2008; Van Peppen *et al.,* 2008; Duncan and Murray, 2012). This is one of many issues that will be explored using qualitative methodology with osteopaths (Chapter 3).

In addition to examining the barriers and facilitators to PROMs, the issues surrounding implementation have been discussed by many studies. Duncan and Murray note that routine outcome measurement has not been implemented into clinical practice despite various initiatives spanning 20 years (Department of Health, 2000; Duncan and Murray, 2012). Different explanations have been proposed for this position, including resistance to innovation and change in

routine practice (Trauer *et al.,* 2006). Boyce *et al*. identified some strategies from their systematic review including engaging professionals in the planning stage of an intervention, ensuring high levels of transparency concerning the reason for data collection, and ensuring adequate training when collecting data (Valier *et al.,* 2014; Boyce *et al.,* 2015). While identifying issues at clinician level are important, it is equally important to look at the wider context within which PROM data collection takes place. This can include organisational reasons, organisational support available, issues of access to the collective ideal and the implications of this, and the views of key opinion holders or leaders within professions (Skeat and Perry, 2008; Van Peppen *et al.,* 2008; Duncan and Murray, 2012; van der Wees *et al.,* 2014). These issues are important not only at the planning stage for an intervention, but become increasingly important for the longer term implementation of an intervention (Deutscher *et al.,* 2008; Grol *et al.,* 2013). Such factors will be discussed in greater depth in Chapter 6.

1.2.13 Technology and healthcare

Osteopathic practice has less effect on and is less affected by advances in new technology than other branches of health care *e.g.* medicine and pharmacology. However, advances in technology have assisted in producing instrumentation that has been helpful in evaluating baseline measurement for patients and being able to assess the effect of changes achieved through treatment. This has been particularly relevant in insurance cases *e.g.* among patients who have suffered cervical and lumbar spine injuries following road traffic accidents. The use of equipment such as the electrogoniometer for measuring peripheral joint range of motion (*e.g.* knee joint, elbow joint), pressure algometers for measuring pain thresholds and muscle tensions, and spinal motion analyser systems which focuses principally on the lumbar spine have allowed objective measures of movement and muscle tension to be recorded (Bronner *et al.,* 2010; Finocchietti *et al.,* 2015; Robert-Lachaine *et al.,* 2016).

However, the ever increasing emergence of new technology has potentially great uses for data collection in healthcare. The information technology (IT) journey began in 1613 when the word “computer” was first recorded (Halacy, 1970). It

referred to a person who carried out calculations, and it was not until the end of the 19th Century that it took on the contemporary meaning of computer. The concept of a programmable computer was associated with Charles Babbage, and his work was developed further by his son, Henry Babbage (Randell *et al.,* 1995). The first fully automatic digital computer, known as the Z3, was developed by Zuse in 1941 (Eckert and Mauchly, 1947; Zuse, 1984). Technology developed during World War II based on the work of Alan Turing and his proposed device of a "single computing machine" underpinned a stored mechanical computer memory for all programme instructions but distinct from today's concept of computer memory (Turing, 1937). Developments in electronics revolutionised computer development, and the advent of the integrated circuit was a notable milestone (Taylor, 1984). The use of integrated circuits within a transistor allowed miniaturisation of computers to progress, and mobile computers grew increasingly popular (Eadie, 1968; Barna and Porat, 1976). Integration of computing resources into mobile telephones underpinned the transformation of mobile phones to Smartphones and Tablet computers which have an increasing market share reaching 1959 million devices shipped by the second quarter of 2016 (Gartner, 2016).

Innovation can sometimes be seen as a double-edged sword; when access is increased to a wider population it can sometimes limit access to others. This has been a noted concern about the growth of technology in healthcare, and the increased assumption that everyone has access to some form of electronic communication facility (Longley and Singleton, 2008). In 2015, 78% of households had fixed broadband internet access, compared with only 57% in 2006, and six in ten respondents (61%) said they personally used their mobile phone to access the internet in 2015 compared with 57% in the first quarter of 2014 (OFCOM research report, 2015). The use of different devices is summarised in Figure 1.8 based on data from the 2013 OFCOM research report.

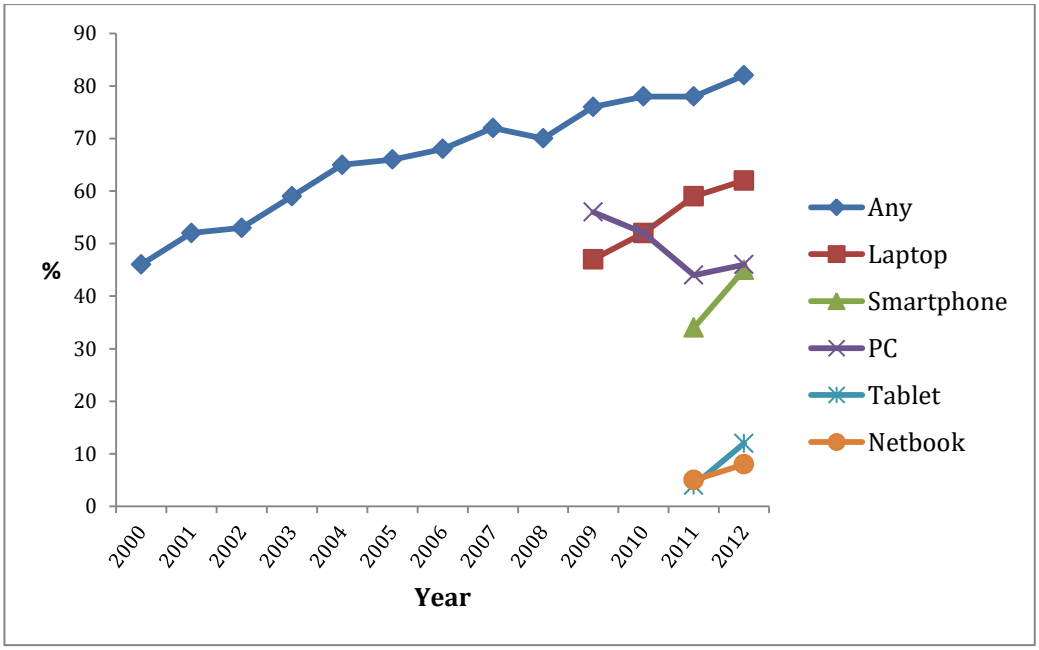


Figure 1.8. Electronic communication use across devices from 2000 to 2012

Despite concerns about the ability to access and use electronic data capture systems in healthcare, there is a growing body of evidence which highlights different systems in use and feedback gathered from users of all ages and health states. A range of different devices have been investigated. Boissy *et al.,* 2006 investigated bar-code scanning as a means of data entry on Personal Digital Assistants (PDAs) when compared to pen-and-paper to allow completion of self-report questionnaires. They identified that while participants found the system enjoyable to use and easy to access there were concerns about the responsiveness of the system which could hinder wider-scale use. Dale and Hagen, 2007 used PDAs in comparison with patient-completed diaries and found that while the PDAs performed better than pen-and-paper in most outcomes, technical malfunction was the biggest hindrance to their use. The issue of physical capability for transferring data has been highlighted in work by Russell *et al.,* 2002. This study compared Internet and paper-based data collection but also investigated the use of two bandwidths *i.e.* ISDN at 128kbit and 17kbit. Bandwidth and speed were found to have no significance on any of the measures suggesting that even home-based bandwidth provision could allow data to be submitted by patients at home if Internet access was available. These speeds are, however, small compared with today's delivery of 5Mb/s. While wider access to technological devices is growing,

a limiting factor of infrastructure development remains in rural areas. The Department for Culture, Media, and Sport (DCMS) reported that public investment of £1.7billion will extend the reach of superfast broadband to 95% of the UK by 2017 (DCMS, 2014). DCMS defines superfast broadband infrastructure as enabling download speeds of at least 24 Mbp/s. The remaining 5% provision remains more challenging and will be provided in different regions (Wales, North Yorkshire, North Lincolnshire, Northern Ireland and Scotland, and Devon and Somerset) by wireless, satellite, and mixed formats including fibre, fixed wireless, sub-loop unbundling (BDUK, 2015).

Other mobile devices have been tested in healthcare settings including touch screen computer systems (Greenwood *et al*., 2006; Salaffi *et al.,* 2009). This technology was found to be a perfectly acceptable option and was not affected by previous experience of computer use. Tablet computers have been tested in some settings. Horng *et al.,* 2012 studied their use in an Emergency Department. This was found to be a feasible option in a busy clinical setting; it was also found to be associated with a reduction in the amount of time clinicians had to log on to a computer. Horng *et al.* make the suggestion that this reduction in time at workstation computer could result in increased availability for patient contact but this association will require further research to support or refute it.

Use of the Internet to support healthcare research continues to grow and develop (Murray, 2007). While a range of different electronic communication devices have been developed to support this initiative, both on terms of clinician and patient use, the equivalence of PROMs between different formats remains unconfirmed. Historically, PROMs have been developed for use in a paper-based format, and while there may be sound reasons for moving to an electronic version of PROMs their validity and reliability when translated to a new format of administration warrants further investigation.

Some work in this area has been undertaken. Lee and Kavanaugh, 2007 identified that computerized versions of their questionnaires produced improved data capture with less ambiguity, less long-term cost, immediate scoring and availability of information, and the ability to collect data more frequently to

monitor patient progress more closely. Saleh *et al.,* 2002 investigated the comparability of the Medical Outcomes Study (MOS) questionnaire, the 36-item Health Survey (SF-36), and the Western Ontario and McMasters University Arthritis Index (WOMAC) when used in either paper format or palm-top computer (Saleh *et al.,* 2002). Few statistically significant differences were identified for the mean, variance, and intra-class correlation coefficient values between the different methods of administration of the questionnaires. However, the internal consistency of the scales was dissimilar highlighting a lack of reliability across modes. This view was challenged by Gwaltney *et al.,* 2008 who undertook a review to investigate the direct comparisons between electronic and paper-and-pencil administration of PROMs. A total of 233 direct comparisons were made and identified that the average mean difference between administration modes averaged 0.2% of the scale range (*e.g.*, 0.02 points on a 10-point scale). In total, 93% were within 5% of the scale range. Among 207 correlation coefficients calculated between paper and computer instruments (typically intraclass correlation coefficients), the average weighted correlation was 0.90; 94% of correlations were at least 0.75. Gwaltney *et al.,* 2008 identified that in four comparisons that evaluated both, the average cross-mode paper-to-computer correlation was almost identical to the within-mode correlation for re-administration of a paper measure (0.88 vs. 0.91).

The most recent study within this area has been undertaken by Bishop *et al.* who tested online and pen-and-paper versions of the Roland Morris Disability Questionnaire (RMDQ) (Bishop *et al.,* 2010). Equivalence at group and individual levels was tested to identify if the different versions of the RMDQ could be used interchangeably. For the study limits of equivalence were pre-defined as 0.5 RMDQ points, the Bland-Altman range was calculated, and participants' comments were examined using content analysis. On analysis of qualitative data, participants identified what they regarded to be unique advantages and disadvantages associated with each version of the RMDQ. However, they confirmed the potential value to be had from offering them the choice of completing the RMDQ online or on paper. The researchers concluded that at both group and individual level the online and paper versions of the RMDQ are equivalent and can be used

interchangeably. Since the trend to electronic data capture appears to be growing, the need to identify equivalence between originally developed versions of PROMs with an electronic equivalent is clear. This information concerning the RMDQ is extremely valuable to researchers and clinicians who wish to use it for electronic data capture. However, the musculoskeletal questionnaire developed for a private practice setting for manual therapists is the Bournemouth Questionnaire, and this currently lacks equivalence testing between electronic and paper versions.

Encouraging the population to complete questionnaires online or using mobile devices requires access, capability, and motivation. The studies examining use of electronic questionnaires are limited. Jenkins *et al.* in their study of patients attending an orthopaedic outpatient clinic reported that 72% of patients reported having internet access (Jenkins *et al.,* 2016). Lack of access, however, was associated with increased age and socioeconomic deprivation. In contrast, Malhotra *et al*. also reporting on patients attending an orthopaedic clinic, found that 85.9% of patients completed an electronic PROM (ePROM) with 50.9% reporting completion at home or work prior to their appointment (Malhotra *et al.,* 2016). A total of 31.5% used a mobile device (Smartphone or Tablet) to complete the ePROM. This concurs with the experience of Bushnell *et al.* in their 2003 asthma study, but there is no information concerning acceptability of use from osteopathic patients (Bushnell *et al.,* 2003).

Age has been suggested as a barrier for ePROMs completion, and the population demographics in the United Kingdom are changing considerably. One of the most notable demographic changes is in relation to age where the size of the population aged 60 years and over is overtaking that aged 16 and under (Figure 1.9).

The percentage of the population living longer has translated into an older workforce as life expectancy has increased to 81 years albeit with regional variation *e.g.* 82.4 years in South East England and 79.1 years in Scotland (Newton *et al.,* 2015; Public Health England, 2016). It is estimated there are 1.13 million

people aged 65 and over in employment, and this can add to the musculoskeletal burden that occurs with increasing age (Fejer and Ruhe, 2012).

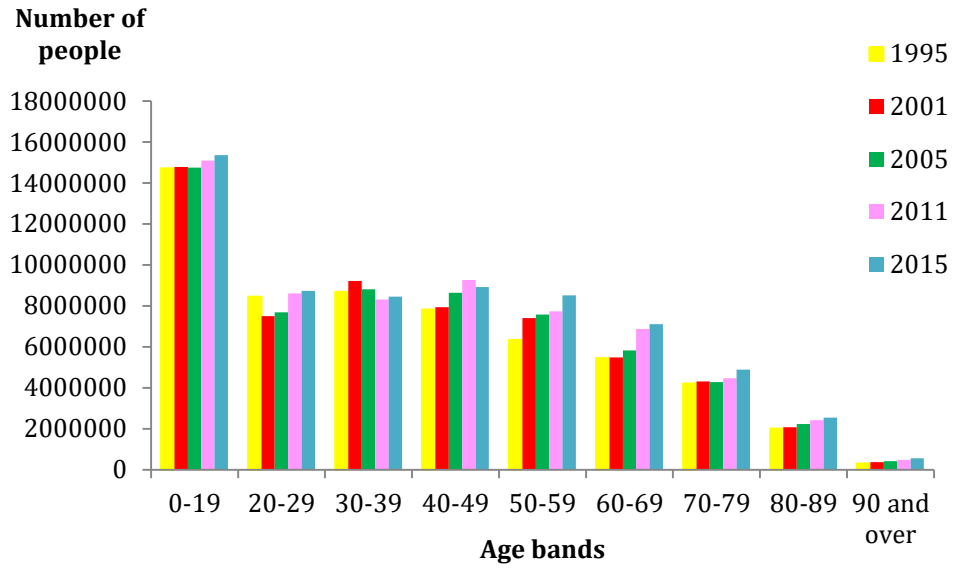


Figure 1.9. Changing age demographic for the period 1995 to 2015 (Based on ONS data, 2016)

Since many of these patients may present at osteopathic practices, it is important to identify in depth patients' views on collecting information about their management, and their preference or lack of it for using electronic data capture (Fawkes *et al.,* 2014b). Qualitative methodologies allow greater exploration of individuals' views on specific topic areas: this will be described in greater depth in Chapter 2.

Data collection and information governance

Data collection and the uses of such data have been topical in the past two years, especially since the introduction of political and legislative changes (Health and Social Care Act, 2012; NHS England, 2014). Considerable changes have been introduced in the NHS since the passing of the Health and Social Care Act (2012) as shown in Figure 1.4. This has not only created new structures in the way the NHS is accessed by patients but it has also created changes in the way the NHS functions; this is notable concerning data handling. This has been summarised in Figure 1.10 by Phil Booth of the patient data pressure group MedConfidential.

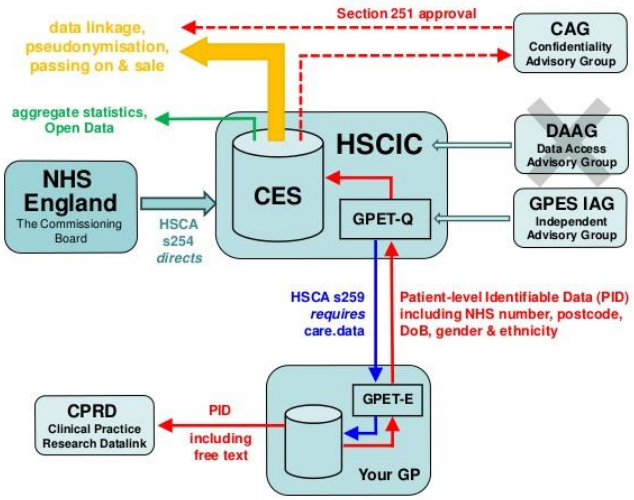


Figure 1.10 Structures in Current NHS Data Management. Source: Booth P. Med Confidential, 2014. Reproduced with permission.

The passing of the Care Act (2014) has made explicit the desire for patients' data to be accessed by research teams, commercial organisations *etc.* with the intention of pooling data to look at trends which could inform healthcare management strategies. Inevitably with new innovation there are opportunities but there are also potential challenges. The lessons from this debate are diverse but important when listening to the concerns of both patients and clinicians, and the advocacy of groups who see the value in sharing large data sets (NHS England, 2014; MedConfidential, 2014). Under the Health and Social Care Act (2012), NHS England has the power to direct the Health and Social Care Information Centre (HSCIC) to collect information from all providers of NHS care, including general practices. The specification of the data to be extracted by GP practices were considered by the Joint GP IT Committee of the British Medical Association and the Royal College of General Practitioners, as well as an independent advisory group (NHS England – Care Data, 2014). In response to concerns voiced by a range of different groups NHS England reported that "*the Independent Information Governance Oversight Panel (IIGOP), chaired by Dame Fiona Caldicott, has agreed to advise the Care Data Programme Board and Senior Responsible Owner on the first phase of the implementation of the programme in its role advising, challenging and reporting on the state of information governance across the health and care system in England*" (IIGOP, 2014; NHS England, 2014).

Although the underlying principle of collecting patient data is sound, there are a number of caveats to be observed to ensure that data are collected in a manner that is both ethical, and with a clear and explicit purpose in mind. Standards for data collection are enshrined in Law (Data Protection Act, 1998), and have been translated into key principles under the work of the Department of Health and Dame Fiona Caldicott, as mentioned above, to form the Caldicott principles (Caldicott, 2013). The findings of the enquiry into the Mid-Staffordshire NHS Foundation Trust (Mid-Staffs Enquiry, 2014) outlined in the Francis report (Francis, 2014) have highlighted the importance of making explicit information about care while balancing that with clear levels of safe-guarding of patient sensitive information: ongoing data collection contributes to making care delivery and outcomes explicit to patients, their carers, and those professionals delivering care. The six Caldicott principles are shown in Table 1.1.

When collecting data in any setting whether in the NHS or private practices it is important to be mindful of the Caldicott Principles. The minimum data necessary should be collected, practitioners should be quite clear about their responsibilities concerning use of the data, and data collection must be carried out with knowledge of the law and complying with the law. Patient data are privileged information and have to be treated with respect. The Caldicott principles are a means of enshrining that respect within a framework. The Data Protection Act (1998) gives rights to patients while requiring individuals who record and use their data to act according to certain responsibilities and standards. There are clear requirements concerning transfer and disclosure of information. Within larger organisations infrastructure and processes exist under the responsibility of a Caldicott Guardian. Although most osteopathic practices do not have the size of capacity to necessitate such frameworks, the spirit of the requirements needs to be observed. The requirements of the Data Protection Act (1998) come with legal responsibilities, and the Caldicott principles bestow an obligation to support the trust patients place in healthcare professionals.

Table 1.1 The Caldicott Principles (Department of Health, 1977).

1. Justify the purpose(s) of using confidential information

Every proposed use or transfer of patient-identifiable information within or from an organisation should be clearly defined and scrutinised, with continuing uses regularly reviewed, by an appropriate guardian.

2. Do not use patient-identifiable information unless it is absolutely necessary

Patient-identifiable information items should not be included unless it is essential for the specified purpose(s) of that flow. The need for patients to be identified should be considered at each stage of satisfying the purpose(s).

3. Use the minimum necessary patient-identifiable information that is required

Where use of the patient-identifiable is considered to be essential, the inclusion of each individual item of information should be considered and justified so that the minimum amount of identifiable information is transferred or accessible as is necessary for a given function to be carried out.

4. Access to patient-identifiable information should be on a strict need-to-know basis

Only those individuals who need access to patient-identifiable information should have access to it, and they should only have access to the information items that they need to see. This may mean introducing access controls or splitting information flows where one information flow is used for several purposes.

5. Everyone with access to patient-identifiable information should be aware of their responsibilities

Action should be taken to ensure that those handling patient-identifiable information - both clinical and non-clinical staff - are made fully aware of their responsibilities and obligations to respect patient confidentiality.

6. Understand and comply with the law

Every use of patient-identifiable information must be lawful. Someone in each organisation handling patient information should be responsible for ensuring that the organisation complies with the legal requirements.

Patients within osteopathic settings disclose sensitive personal information and respecting this in accordance with Caldicott principles engenders trust. A range of different topics relevant to osteopathic practice specifically, and healthcare generally have been discussed in Section 1.2. Gaps in the literature have been identified, and I will describe in the next section how these will be translated into the content for this thesis.

1.3 Outline of the thesis

In the following chapters, I will begin to address my objectives by employing a qualitative research approach to identify the views of patients about the concept of data collection in clinical practice, and their views concerning three specific outcome measures (Objective II). This will be followed in Chapter 3 by undertaking further qualitative work to identify osteopaths' views concerning PROM data collection (Objective III), and the views (and experiences where possible) of physiotherapists and chiropractors concerning PROM use in clinical practice (Objective IV).

In Chapter 4, I will undertake a systematic review of the Bournemouth Questionnaire (Objective V). This information combined with the findings of the qualitative chapters (2 and 3) will inform development of the content of a PROM data collection app (Objective VI). The initial version of the app will be pilot tested and the findings of the pilot will be examined to inform any necessary changes for the app (Objectives VII, VIII and IX). This stage of the thesis will be described in Chapter 5.

On completion of the pilot testing, the content of the app will be refined and it will be introduced into day-to-day clinical practice with volunteer osteopaths (Objective X). This stage and early findings of the implementation stage will be described in Chapter 6.

In the final discussion, Chapter 7, I will summarise the issues which have arisen during the study, and consider the implications of the study on current PROM use

and potential electronic PROM use. Areas for future research will be discussed briefly, and I will draw on my experiences to highlight areas of strength and weakness in my research generally.

2

A qualitative study – investigating patients' views concerning the use of Patient Reported Outcome Measures (PROMs).

2.1 Introduction

The role of patients in healthcare has changed considerably during the past 40 years. No longer are they regarded as the grateful recipient of care, but instead as a key part of the therapeutic process (Bobo *et al.,* 1991). At the same time, the number of patients who are living longer, and with multiple co-morbidities is increasing so their involvement in managing their health should be welcomed (Salisbury *et al.,* 2011; Barnett *et al.,* 2012). With an increasing focus on outcomes of care in both short-term and long-term conditions, it is important that those outcomes should encompass items that are important to the patient in addition to clinicians, managers, and funders (Murphy *et al.,* 2015).

While the evaluation of patients' symptoms can be identified using a range of clinical measurements, and through questionnaires in research studies, there can

be a disconnect between those symptoms and activities which are important to the patient, and what is perceived by researchers and clinicians to be important to patients (Pannucci and Wilkins, 2010; Chiarotto *et al.,* 2015). Patient involvement in their management and research has been improved considerably with a range of different initiatives, but there is still room for further change (Kemshall and Littlechild, 2000; Ellis *et al.,* 2014).

2.1.1 Background

The last decade of the twentieth century marked a period of considerable increase in patient involvement in research. Starting as a legal requirement of community development initiatives in the late 1960s, patient involvement has developed as service users have helped to shape public and social policy (Beresford and Croft, 1992; Means and Smith, 1998). This has been accompanied by use of new terminology *e.g.* "user-involvement", "partnership", and "empowerment" (Beresford, 2002). Beresford suggests that "*the emergence of these new movements was assisted by the challenge to traditional paternalistic welfare represented by the political New Right and its rhetorical emphasis on the 'active citizen', and individual rights and choice.*" There may be some truth in this, but the emphasis on patients climbing the "ladder of participation" continues to increase in professional and political agendas, and other attendant initiatives (Arnstein, 1969).

The desire to increase patient engagement and participation in healthcare is often cited as one of the many drivers for the introduction of PROMs in healthcare (Appleby and Devlin, 2010). Improving self-management, enabling greater individualisation of care, promoting better communication, assessing effects of treatment, and influencing clinical management through the monitoring and detection of symptoms are commonly cited also as valuable reasons for their progressive introduction into healthcare (Greenhalgh *et al.,* 2005; Greenhalgh, 2009; Appleby and Devlin, 2010; Dawson *et al.,* 2010; Snyder *et al.,* 2013; Hunter *et al.,* 2015). This is occurring despite some scepticism by some clinicians concerning the validity of PROMs, and antipathy by some clinicians and researchers about increased patient/user engagement (Bream *et al.,* 2009; Snape *et al.,* 2014). Despite the fact that patients play a key role in the healthcare system, their views

concerning new innovations and the management process are less commonly sought.

While the rationale for the introduction of PROMs into clinical practice is sound from a clinician's, manager's, or commissioner's viewpoint, the perspective of the patient is a vital part of their successful implementation. The introduction of PROMs into the NHS continues to grow, notably through the national PROMs programme which is beginning to publish data, and through required completion prior to outpatient appointments (NHS Digital, 2016). There is, however, less documentary evidence of their use in primary care in the NHS and in private manual therapy practice settings (Murphy *et al.,* 2015). The availability of suitable PROMs for primary care has been cited for this as many patients are consulting with multiple conditions (Fitzpatrick *et al*., 1998). The same can be said for osteopathy where patients present with multiple conditions, albeit with some presenting conditions being more frequently reported than others (Burton, 1981; Pringle and Tyreman, 1993; Hinkley and Drysdale, 1995; McIlwraith, 2003; Fawkes *et al.,* 2012). This does provide some scope to identify a range of PROMs potentially suitable for osteopathic practice since low back pain and neck pain were the most frequently presented symptoms.

While the selection of a particular PROM and other measures of outcome is important, of equal importance are considerations about how and where the PROM should be completed, how it should be administered, over what duration to provide data without this becoming burdensome, and what information should be made available to the patient during completion of the PROM or at the end of the therapeutic management window (Hildon *et al.,* 2012a; Hildon *et al.,* 2012b; Hildon *et al.,* 2012c; Ryan *et al.,* 2016).

The use of PROMs in selected populations with symptomatic or functional problems appears somewhat challenging. However, even more challenging are settings where the patient population is both unselected and presenting with a range of diverse symptoms, conditions, and reasons for their consultation. This is the perplexing situation faced in primary care. While generic PROMs exist, for

example the EQ5-D, and the SF-36, some commentators regard these PROMs to be lacking sensitivity to change for patients in primary care who may present with neither symptomatic nor functional problems (McDowell, 2006; Binns *et al.,* 2007). Starfield described primary care as providing a range of functions including the provision of first contact to resolve symptoms or requiring reassurance that symptoms are not indicative of more serious illness, the provision of ongoing care for patients with multiple comorbidities, the provision of advice, or signposting to other services Starfield, 1979; Murphy *et al.,* 2014).

The short-term and longer-term nature of patient management in primary care presents challenges for PROMs' use. For example consultations can involve the resolution of current or immediate symptoms or they may involve the episodic management of symptoms whose outcome may not be known for some years hence (Valderas *et al.,* 2012). To meet this challenge, Murphy et al investigated patients' and clinicians' views on the most important outcomes arising from primary care consultations (Murphy *et al.,* 2015). This qualitative work involving 30 patients and eight clinicians in the South West of England identified three main areas of importance relating to outcome. These included health empowerment (patients' understanding of their illness; ability to self-care and stay healthy; agreeing and adhering to a patient-clinician shared plan; confidence in seeking healthcare; and access to support), health status (reduction of symptoms; and reducing the impact of symptoms on patients' lives), and health perceptions (patients' satisfaction with their health; health concerns; and confidence in their future health).

This qualitative study formed the basis of a Delphi consensus study involving patient, clinician, and academic panels to rate outcomes which were relevant to health, relevant to primary care, and detectable by patients (Murphy et al., 2016a). A 27-item instrument known as the Primary Care Outcomes Questionnaire (PCOQ) was created from the Delphi study, and has now undergone quantitative evaluation involving 602 completed questionnaires from primary care patients. This questionnaire is now available for use and evaluated the outcomes of primary

care management and interventions from a patient's perspective (Murphy *et al.,* 2016b).

In contrast to primary care, data collection NHS outpatient settings takes place commonly on site, and prior notification of need for support in terms of translation is identified and provided (Ritchie, 2015 – personal communication). Data collection tends to be paper-based in many locations, although an increased move to make PROM completion part of electronic patient records is developing as the NHS moves towards a paperless status (Smith *et al.,* 2005; Hunt, 2013; Ritchie, 2015). This has been recommended by 2018, and is included as part of government policy on NHS efficiency from 2010-2015 (UK Government, 2010; Hunt, 2013). The value of showing patients their previous scores on PROMs or other measures during a consultation is much-debated: advocates note that it increases patient engagement and supports communication with clinicians, but detractors raise the issue of panel-conditioning (Ryan *et al.,* 2015; Underwood *et al.,* 2006a). Notwithstanding the challenges of gathering baseline and follow-up data, the content and format of how data should be fed back to patients and clinicians has had limited examination. While there is a paucity of literature in this area, and what is available is heterogeneous, some recommendations have been made concerning the most useful way to communicate PROM findings to patients and clinicians (Hildon *et al.,* 2012a; Hildon *et al.,* 2012b). Much of the available literature on communication findings focuses on the communication of risk to patients, but healthcare can benefit from outputs by business literature in terms of identifying useful visual presentation of data (Edwards *et al.,* 2006; Schapira *et al.,* 2006; Gerteis *et al.,* 2007; Fasolo *et al.,* 2010).

While the data concerning the communication of risk has undergone focussed research in osteopathy in recent years, there has been no research examining the communication of findings from PROM data to patients (Gibbons and Tehan, 2006; Froud *et al.,* 2008; Carnes *et al.,* 2009; Carnes *et al.,* 2010 Leach *et al.,* 2011; Vogel *et al.,* 2013). This aspect of patient care and the value patients place on the usefulness of PROM data collection in osteopathic practice will be among the issues considered in this chapter.

2.1.2 Overview of qualitative research methods

Qualitative research offers a broad methodological approach to enquiry and contains many different methods which have been utilised by social and natural sciences, business, and market research (Denzin and Lincoln, 2005). While qualitative research methods have been used in the social sciences for a long time, historically they have been used less frequently in the biomedical or natural sciences. More recently researchers in healthcare have recognised the potential of qualitative research when used to complement quantitative methods. At the root of some of the antipathy to qualitative research lies a misunderstanding about the nature of qualitative research which has been labelled as "unscientific" (Pope and Mays, 2006; Walsh and Downe, 2005). Other misunderstandings have labelled qualitative research as:

- Subjective (and inherently biased);
- Difficult to replicate;
- Amounting to little more than anecdote, conjecture, or personal impression;
- Not generalisable to the wider population as it involves such small samples of participants.

Notwithstanding the above criticisms, qualitative research has much to offer to healthcare research. A less contentious definition which describes the potential of qualitative research was developed by Strauss and Corbin (Strauss and Corbin, 1998); they described qualitative research as any type of research that produces findings not arrived at by statistical procedures, or other means of quantification. While other definitions exist, Ritchie and Lewis identified key characteristics of qualitative research (Ritchie and Lewis, 2009). These include

- "Aims which are directed to both an in-depth understanding and interpretation of the social world of research participants by learning about participants' social and material circumstances, their experiences, histories, and perspectives";
- "Samples that are purposively selected on the basis of specific criteria, and small in scale";

- "Data collection methods which usually involve close contact between the researcher and participants, and allows the opportunity for emergent issues to be explored";
- "Data which are very information rich and extensive";
- "Analysis which is open to emergent concepts and ideas which may produce detailed description and classification, identify patterns of association, or develop typologies and explanations";
- "Outputs which tend to focus on the interpretation of social meaning through mapping and 're-presenting' the social world of research participants". This can have the added benefit of informing subsequent qualitative research through identifying and then allowing exploration of a range of opinions.

Qualitative research essentially attempts to answer many of the "why" questions, which arise in the research process, and in some cases the "how" related to decision-making. Increasingly it acts to complement quantitative research findings, or is conducted to inform the development of quantitative data collection *e.g.* as nested studies, by considering the totality of a situation (Lewin *et al.,* 2009). Silverman asserted that

"Dependence on purely quantitative methods may neglect the social and cultural construction of the variables which quantitative research seeks to correlate" (Silverman, 2011).

Within qualitative research, a range of different theoretical positions exist including:

- Phenomenology;
- Ethnomethodology;
- Grounded theory;
- Symbolic interactionism;
- Interpretivism;
- Constructivism.

Each methodology will be considered briefly in turn:

Phenomenology

Phenomenologists "reject" statistical and quantitative explanations for human behaviour as they feel such approaches cannot produce a causal explanation for such behaviour (Marton, 1986; Harris, 2016). Phenomenologists regard human beings as being capable of making sense of the world by applying their own unique meaning and classifications to the world around them to make sense of their own social world.

Ethnomethodology

This is based on an American sociological perspective. It is concerned with the study of society and how a phenomenological perspective can be used to explain the meanings and classifications are responded to by social groups and wider society (Atkinson, 1988; Greiffenhagen *et al.,* 2015).

Ethnography

Ethnography is concerned with the study of social interactions, and the behaviours and perceptions that are displayed within groups of individuals, organisational cultures, and other social teams. It developed from anthropological studies in the early 1900s and is characterised by exploring the nature of social phenomena, often using unstructured data, and studying small numbers of groups or "cases" Hammersley and Atkinson, 1995; Reeves *et al.,* 2008).

Grounded theory

Grounded theory was developed by Glaser and Strauss (Glaser and Strauss, 1967). Its main focus is the generation of theories which allow the development of a higher level understanding of social phenomena, essentially exploring how this is "grounded" in analysis of data. Grounded theory is used to study social interactions or experiences, and aims to explain a process. To achieve this it employs an iterative study design, theoretical (purposive) sampling, and system of analysis (Kennedy and Lingard, 2006; Lingard *et al.,* 2008).

Symbolic interactionism

Symbolic interactionism is linked conceptually very closely with grounded theory, although more recently attempts have been made to try to disentangle this connection (Handberg *et al.,* 2015). It is underpinned by three basic tenets including that human beings act towards things, based on the meanings they ascribe to those things, the meaning ascribed to things arises from social interaction with other human beings, and meanings are handled and modified through an interpretive process (Blumer, 1969).

Interpretivism

Interpretivists believe that qualitative research is one of the fundamental tools of sociology in its aims to understand social actions by being able to understand the meanings and motives underpinning behaviour. Interpretivism is often linked to the thoughts of Max Weber (1864-1920) who suggests that in the human sciences we are concerned with *Verstehen* (understanding) in comparison to *Erklaren* (explaining), essentially focussing on process instead of 'facts' (Weber, 1991). Through this method, the interpretivist looks to interpret the social world through culturally and historically derived means (Hughes, 2016).

Constructivism

Constructivists share their investigation with their research subject. They do not tend to begin with a theory but develop a theory or pattern of meaning through an inductive process as the research process progresses (Cohen and Manion, 2011). They tend to rely on the participants' views of the research being studied while recognising the impact of the research on their own background and/or experience (Cresswell, 2003; Johnson and Onwuegbuzie, 2004.

Within each of these theoretical approaches, a range of different methodologies exist to undertake qualitative research. These include:

- Individual interviews;
- Focus groups;

- Observations;
- Action research.

All of the different theoretical perspectives and methodological perspectives have their imitations. One notable limitation is that the data are collected by a researcher, analysed by other researchers, and become increasingly remote from the initial raw data collected (Weber, 1946; Coomber, 1997; Pope and Mays, 2009). The researcher is regarded as "an instrument in the research process" and they should ensure their accounts are credible (Pyett, 2003). The theoretical frameworks with which analysis will be undertaken will move the data further from its original state also. However, recognising that these limitations are part of the process, the researcher should be aware of and reflect upon how these influences will affect their analysis, and the overall findings of their research, *i.e*. be reflexive. The notion of reflexivity has been increasingly documented as qualitative research has grown, its importance has been emphasised by many authors including Bourdieu who describes its importance:

"*the subjective relation to the object – which, when not taken into account, and when it orientates choices of object, method, etc., is one of the most powerful factors of error [in research]"* (Bourdieu, 2004 cited in Rae and Green, 2016).

The focus on reflexivity is echoed by Finlay, and Rae and Green propose a model overlaying the concepts of both authors to promote high levels of reflexivity in qualitative research (Finlay, 2002; Rae and Green, 2016). While reflexivity is an important part of adding robustness to qualitative research, this quality can be enhanced further by the separate, and preferably independent, researcher evaluating the raw data to ensure consistency in interpretation of the data in a process described as "triangulation". It has been defined by Cohen and Manion as an "attempt to map out, or explain more fully, the richness and complexity of human behaviour by studying it from more than one standpoint" (Cohen and Manion, 2000). Triangulation can involve also involve different people and processes, and Denzin identified four types of basic triangulation. These included data triangulation (involving different time points and people), investigational

triangulation (involving one or more researchers), theory triangulation (involving more than one theoretical standpoint for data interpretation), and methodological triangulation (using different methodological approaches to gather data) (Denzin, 1978). The data yielded by qualitative research methods complement and inform quantitative methods, and the attention to ensuring robustness in data collection and analysis has facilitated its increased use in medical research (Underwood *et al.,* 2006a).

I will now expand upon the aims for this qualitative study.

2.1.3 Aims of the study

The aims of this study were:

- to explore patients' views on the general idea of routine data collection in clinical practice;
- to explore their preferences for particular types of data collection media;
- to explore their views on particular PROMs identified prior to treatment.

2.2 Methods

I used a qualitative phenomenological approach because it allowed me to explore patients' views about PROMs in depth: a quantitative approach would have been more restricted in allowing patients to "sing their songs" (Catlett, 1937). A range of guidelines exist to advise authors about their content when reporting particular types of studies. Guidelines created by Tong *et al.,* 2007 have been used in the methodological description of this study.

Individual interviews were used based on a discovery-orientated approach to explore in-depth patients' views on the use of PROMs in clinical practice, focusing on specific PROMs and their suitability for assessing their symptoms, and different methods of data capture *i.e.* paper, web-based, or using an app. This enabled exploration of patients' views on what symptoms are important to them when

considering change, and the most meaningful way to describe such symptoms *e.g.* whether through text statements or through numerical rating evaluation. The interview approach allowed me to follow up interesting comments or remarks with other questions for clarification or expansion. This approach was favoured to a focus group approach for a variety of reasons, but I was particularly keen in this study to avoid the type of dynamics which can arise when certain group members are more assertive or compelling than others in making their views known. This can influence other group members and introduce a form of "group bias" (McGee, 1999).

Prior to beginning any form of research, it is important to identify initially if the intended course of action actually is research, and then ensure that the rights and wellbeing of patients are protected throughout (NHS R&D Forum, 2014; World Medical Association, 2013). Once this had taken place, the protocol development began, and attention was drawn to considering ethical concerns, and the manner in which the study was delivered in accordance with good research practice (Department of Health, 2005; MRC, 2012). This will be explored in the next section.

2.2.1 Ethics and research governance

This research study used individual interviews with research participants. The aim of such an approach is to ask in-depth questions, and respond to answers by pursuing particular topics that may arise. In some instances patients may become upset at the content of questions, and this must be reflected upon by the researcher prior to the interview taking place, and a strategy envisaged for how to deal with such situations. When reviewing my topic guide, I didn't consider that any of my topics or questions would be likely to cause distress to participants. However, to guard against this, prior to the interview taking place, participants were sent a participant information sheet (PIS) outlining the reason for the study aiming to answer any questions about the process, and how the interview data might be used. Participants' right to terminate the interview or withdraw from the study at any time without having to give any explanation was also made explicit.

Patients were informed that their interviews would be audio-recorded, transcribed verbatim, and unattributed comments might be used in future publications. Any information that could potentially identify the participant would be removed from the transcription also. Quotations would be identified with a code generated during the analysis process and would be known to the research team only. Participants were also informed that their audio-recordings and transcribed data would be stored securely for 20 years, and would be accessible only to the study team during that period.

A consent form was included with the PIS, and participants were asked to sign and return the consent form to me before the interview took place to ensure they fully understood the study and were willing to participate. Copies of the participant information sheet and the consent form are included in Appendix 2.1and Appendix 2.2 respectively. The study protocol was prepared in accordance with Good Clinical Practice, reviewed internally, and ethics approval was obtained from the research ethics committee at Queen Mary University of London (REF: QMREC2013/57).

2.2.2 The sample and recruitment

One of my main aims in this PhD was to create a data collection facility suitable for osteopathic practice. Consequently the sample in this study was drawn from the population of patients attending private practices for osteopathic treatment. This sample was identified by purposive sampling of osteopathic practices. Patients who received osteopathic care funded by the NHS, or were treated in NHS settings were excluded from the sampling due to the limits of my ethics permission. Participants were recruited from a variety of sources including recruitment through posters displayed by practising osteopaths throughout the UK, information sheets available in practices, and through the Patient and Public Partnership Group created by the profession's regulator, the General Osteopathic Council (http://www.osteopathy.org.uk/about/our-work/Patient--Public-Partnership-Group/). Recruitment took place over a period of eight months using a purposive sampling framework.

A purposive sample is a non-representative subset of some larger population, and is constructed to serve a very specific need or purpose. A qualitative researcher may have a specific group in mind and will attempt to focus on the target group, interviewing whomever is available (Ritchie and Lewis, 2009). Purposive sampling is most commonly used in qualitative research in comparison to other forms of sampling used in quantitative research (Altman, 1999). Earlier studies have identified demographic information concerning the population attending private practices for osteopathic treatment (Fawkes *et al.,* 2014b). The potential impact that the different criteria could have on views concerning data collection, and how data might be collected were considered when planning this qualitative study.

Semi-structured interviews were used based on a topic guide, and this is described in the next section.

2.2.3 The Topic Guide

The content for the topic guide was developed following review of the literature concerning PROMs, and involving my supervisors (DC and RF). Brainstorming aimed to identify the issues that are important to patients when their response to treatment is being assessed. The Topic Guide contained a set of themes and broad questions to explore during the interview process. The key areas for investigation were:

- to understand how participants felt about data collection in practice and if they had any concerns;
- to explore their views about three different PROMs and whether they thought the PROMs could capture information which was relevant to them and how their symptoms affected their daily life;
- to identify whether participants had any preference for completing PROMs using a particular format and whether any support would be needed to facilitate their preference(s).

The main topic headings are shown in Table 2.1; the full topic guide is shown in Appendix 2.3.

Table 2.1. Topic guide

Section	Purpose
Introduction	*To introduce the interviewer, thank the interviewee for agreeing to participate, confirm that the interviewee is happy for the interview to be recorded, and check that there are any questions about the interview not covered in the participant information sheet.*
Collecting data in practice	*To introduce the idea of practice-based data collection into clinical practice, and exploring interviewees' views on the acceptability of this, the time they would be willing to spend upon this, and the preferred method for doing this. To enquire about any support necessary to use electronic data capture.*
Concerns/ issues about collecting data	*Interviewees were asked if they had any concern about data being collected in practice per se, and especially if the data were being submitted to a third party in the form of a university-based researcher. Enquiries were made also about where participants would be willing to complete questionnaires and whether they would prefer to do this at home or within a practice setting.*
Helpful information	*Enquiries were made about the type of information interviewees thought would be useful to receive if data were collected in their won practice and nationally. The manner in which this information should be made available was also explored.*
Content of PROMs	*Discussion of a selection of three different PROMs. To explore the usefulness of the questions in trying to capture data which patients felt were relevant from their experience of symptoms. Questions that were missing or superfluous were also explored.*
Other relevant issues	*To enquire about other aspects of the therapeutic process that interviewees though were important which were not captured in the PROMs but they felt were important in terms of data collection.*
Close	*To thank the interviewee for their time. To explore if interviewees wanted to receive information about the progress of the project, and the format in which they would like to receive information.*

2.2.4 Piloting

Once the draft topic guide had been created, a pilot telephone interview was undertaken with a colleague. I wanted to ensure that all questions felt relevant, and that they were not ambiguous. I also sought feedback on how the interviewee felt the process went and whether different aspects felt either too leading or too hesitant. This allowed me to work on different areas of my interview technique where weaknesses had been perceived.

2.2.5 Participant interviews

Telephone interviews were arranged at a time convenient to participants, and lasted between 30 and 45 minutes. They were tape-recorded with the knowledge and consent of the participants. Additional field notes were taken to record any significant events or comments associated with the interviews. Telephone interviews were used for logistic reasons and convenience to try and gain greater geographical spread and variation among participants. Training in qualitative interview techniques was undertaken by me at Oxford University prior to beginning this study. All participants gave signed consent prior to taking part in the interviews.

The content of the interview data was assessed after each interview to identify if any refinements were required in the topic guide to phrase a question slightly differently to add clarity or to add some further dimensions to improve the depth of the response. Any changes and dates were recorded in field notes. A sample size of approximately 15 participants was envisaged for this study to allow inclusion of a range of participants with characteristics identified in Table 2.1. This was based on the timescale to complete this study within the limits of the PhD, and on the findings of a similar qualitative study (Carnes, 2006). Recommended or suggested sample sizes in qualitative research vary from study to study (Guest *et al.,* 2006; Mason, 2010). Thomson explored this in a review of 50 studies which had used grounded theory: samples ranged from 5 to 350 participants (Thomson, 2004). Authors provide guidance on sample size, for example Creswell suggests 20-30 participants, and Morse over 30 (Creswell, 1998; Morse, 1994). In the review by Thomson, 34% of studies used between 20-30 participants, and 22%

used over 30 participants (Thomson, 2004; Mason, 2010). Ritchie and Lewis suggest that key considerations should guide sample sizes including whether new evidence is being generated by each successive interview since "*there is therefore a point of diminishing return where increasing the sample size no longer contributes new evidence*" which is termed 'data saturation' (Ritchie and Lewis, 2012). Additionally, the number of interviews is not the province of qualitative research which places more focus on the depth and richness of the detail in the data collected. Ritchie and Lewis further recommend other items for consideration including the heterogeneity of the population, the qualitative research method, and the number of selection criteria (Ritchie and Lewis, 2012).

2.2.6 Analytical approach

All of the interviews were transcribed verbatim by me. This allowed me to become immersed in the data over time, and also allowed me to develop the topic guide further where different issues arose with participants. All transcripts were anonymised, and page and line numbers were inserted to ease data handling and make identification of references easier when undertaking analysis. Participants were asked to check their transcripts for accuracy to ensure that the content reflected their views in a process termed "member checking". This provided the opportunity also for participants to withdraw any statements with which they felt uncomfortable in hindsight (Guba, 1981).

A range of different approaches can be utilised to analyse qualitative data but I used the "Framework approach". This is a matrix-driven approach developed by the National Centre for Social Research (NatCen) employing a semi-structured interview approach allowing flexibility to explore evolving issues (Ritchie and Lewis, 2012). Pope *et al*. describe five stages of data analysis when using the Framework approach. These include:

- "Familiarisation – immersion in the raw data";
- "Identifying a thematic framework – identifying all key issues, concepts and themes by which the data can be examined and referenced";

- "Indexing – applying the thematic framework or index to all the data by annotating the transcripts with numerical codes from the index";
- "Charting – re-arranging the data according to the appropriate part of the thematic framework to which they relate and forming charts";
- "Interpretation – using the charts to define concepts, map the range and nature of phenomena, creating typologies, and finding associations between the themes with a view to providing explanations for the findings";
- "Creation of matrices – to allow within and between case examination" (Pope *et al.,* 2001).

This approach was helpful and suitable for this type of analysis allowing the content of the transcripts to be examined, and emergent themes to be detected by two independent researchers (CF† and DC*).

The initial phase of this analysis involved reading and re-reading the text of the transcripts to facilitate the immersion process. Any field notes taken during the interviews were examined also to add context to the transcripts. After familiarisation, the transcripts were assessed to see if any themes and topics began to emerge from the interviews. Emerging topics were then placed in categories, and then further placed under common themes.

Organising and managing the data

Interview transcripts were initially allocated a reference code number to preserve the anonymity of participants. Transcript pages were numbered, the process of indexing then took place where each line of the interview transcripts was numbered and used as a reference. Emerging topics were identified and ordered into themes; the process of indexing allows the text to be apportioned to the topic or theme in a systematic manner.

The references were then organised into a framework or a grid. Themes and sub themes (topics) were organised into columns. Participant information was organised in rows so that the data could be viewed along themes or by

† C.Fawkes; * Dr Dawn Carnes

interviewees with specific characteristics involving, for example, age, ethnicity or any other characteristic of interest. This enabled the development of theories about issues associated with PROMs and data collection, and how they related to individual participant characteristics.

Microsoft Excel was used to organise and manage the data generated from interview transcripts. Although software packages are available to organise and manage the data *e.g.* Computer Assisted Qualitative Data Analysis Software (CAQDAS), the Excel software functioned well allowing the data to be interrogated where necessary using the search function. It has been used in several other published qualitative studies (Cope, 2014; Cameron and McCall, 2015; Fargnoli *et al.,* 2015).

The first stage of the analysis involved immersion in the data and identification of the emergent themes from the interview transcripts. Independent theme identification took place also by one of my supervisors (DC*). The second stage involved discussing individual ideas concerning themes, and agreeing a draft framework. The draft framework was then tested against a sample of transcripts to identify if the coding framework required any revision. Once the framework was agreed after discussion, quotes were coded under the appropriate themes and sub-themes from emergent topics. This procedure ensured transparency of the analysis process. Any disagreements could have been discussed and resolved by my second supervisor (RF‡), but this proved unnecessary. This decision-making process was documented to provide an audit trail to support transparency within the decision-making process (Guba, 1981).

2.2.7 Coding, charting, and triangulation of data

A sample of transcripts (n=9) was given to an independent assessor to analyse. The independent assessor chosen, Brigid Tucker (BT), is Head of Policy and Communication at the General Osteopathic Council, the professional regulator charged with the care and wellbeing of osteopathic patients. Mindful of the fact

* DC: Dr Dawn Carnes; ‡ RF: Dr Robert Froud

that the two researchers who analysed the data had been in clinical practice as osteopaths, the data analysis by BT was intended to add the perspective of a non-clinician albeit one who is informed about osteopathic care and patient wellbeing. BT was asked to examine the transcript and try to identify any emerging themes and sub-themes. These were then compared with the framework develop by DC and CF to look for areas of consistency and/or disagreement.

Although three more sub-themes had been identified by BT, other areas consistently overlapped. The additional themes were incorporated into the framework.

2.3 Results

2.3.1 Sample and characteristics

I undertook a total of 22 telephone interviews for this qualitative study. The sample consisted of six men and 16 women who were currently receiving osteopathic treatment in a private practice setting. Participants were aged between 37 and 86 years, specific ages were freely given by some participants although age data were sought according to age bands. The full characteristics of interview participants are shown in Table 2.2.

Data concerning ethnicity were collected and the participant sample contained only one participant who was British Asian; the remainder of the sample were White British (n=21), one of whom noted “White Scottish” as their ethnic status. Data were collected also concerning work status.

Table 2.2 Full characteristics of interview participants.

	Sex		Age						Ethnicity		Work status				Region								
	M	F	31-40	41-50	51-60	61-70	71-80	81-90	White British	British Asian	Working full or part time	Retired	At home	Long term sick	Scotland	Wales	London and South East England	South West	West Midlands	North Midlands	Yorkshire	North East	Sussex
Number of partici-pants (n)	6	16	1	2	8	5	4	2	21	1	9	9	2	2	1	1	6	2	5	3	1	1	2

The comparison between the planned and the actual participants recruited is shown in Table 2.3.

Table 2.3. Actual characteristics of study participants

Characteristics	Male	Female
Age		
18-29		
30-39		√
40-49		√
50-59	√	√
60-69	√	√
70-79	√	√
80 and over	√	√
Duration of symptoms		
Acute		
Sub-acute		
Chronic	√	√
Ethnicity		
White British	√	√
White Scottish		√
White other		
Mixed/multiple ethnic groups		
Asian/Asian British		√
Black/African/Caribbean/Black British		
Other ethnic groups		
Work status		
Working (full time/part time)	√	√
Retired	√	√
Long term sick	√	√
Not working		
At home		√
Other		

A range of characteristics were not represented. Male and female representatives from White "other", Mixed/multiple ethnic groups, Black/African/Caribbean/Black British, and "Other ethnic groups" were unrepresented. Male participants from Asian/Asian British backgrounds were also not present among the interview participants.

Participants reported chronic symptoms, and those with acute or subacute symptoms were not represented. Although there was a good variation in the age

of participants, some groups were unrepresented including males and females from 18-29, and males in the 30-39 and 40-49 age ranges.

Geographical variation was good although there were no participants from Northern Ireland, or the North West of England. Work status was varied among participants with equal numbers of participants who were involved in full time or part time work and those who were retired. There was no representation for participants who were unemployed, or "other" *e.g.* students not in paid work.

2.3.2 Emergent themes and sub themes

Within each of the participants' transcripts themes and sub-themes emerged. These are described in Table 2.4. Data were organised into five separate themes describing different aspects of the data collection process, its potential impact on the consultation and practice visit, the factors that might affect the data collection process, and the manner and content in which information should be fed back to participants/patients.

Theme 1. Data collection

In this theme, I identified issues around the content of the PROM and whether patients regarded their content as relevant to them. The format of the PROM and the manner in which data were captured was discussed allowing participants to express their preference or ambivalence for a particular manner in which data should be collected. The importance of other issues about which data were not collected was also raised by some patients; such issues included general experience of treatment, hygiene issues, and empathy. These non-specific effects and environmental issues may be raised as single questions in some surveys but did not appear in the PROMs discussed.

1.1 Relevance of content

Participants stated that some of the items included in the Patient Reported Outcome Measures did not reflect all of the issues that were relevant to them in

Table 2.4 Emerging themes and sub-themes from patient participants' data

Themes	Sub-themes				
1. Data collection	1.1 Relevance of content	1.2 Potential to measure change	1.3 Timeliness for completion	1.4 Variety of formats (words and numbers)	1.5 Technological issues in completion modes
2. Data protection	2.1 Choice of participation	2.2 Data protection	2.3 Use of data	2.4 Confidentiality of technology	
3. Purpose of PROMs	3.1 Clear statement of purpose of data collection	3.2 Feedback within consultation	3.3 Measuring effectiveness of treatment	3.4 Consideration/reflection on other issues	
4. Motivation	4.1 Information about practices (types of problems treated, *etc.*)	4.2 Widening patient access/choice	4.3 Information for other healthcare professionals		
5. Dissemination and feedback of information	5.1 Information overload	5.2 Variety of dissemination formats	5.3 Clinical versus research purpose		

their experience of pain and disability. Some items listed were superfluous and others were absent.

"I might actually say something at the time like why is that question there…"(P1, page 3, line 119).

"I think again it's sort of the lifestyle stuff and I don't think that there's anything in here about the… I can't do my exercise class because my condition is so bad kind of thing". (P1, page 5, lines 173-175).

"The one thing these questionnaires never capture is the sheer frustration associated with back pain. I find it a real bind having to ask for help". (P4, page 3, lines 119-121).

"I'm often aware of the effect that my condition has on my family, and there is nothing in these things that touches on that". (P20, page 6, lines 191-192).

"The one thing these didn't ask was about the experience of treatment. I've been to some places where they answer the phone and go out and answer the door and ... it kind of intrudes. This is important stuff and I think that should be asked "(P22, page 7, lines 200-203).

1.2 Potential to measure change

The ability of whether PROMs could measure change was discussed. The view that some of the statements were aimed at populations with considerably greater levels of disability was expressed by several participants.

"Well I just thought it was for people who were way worse than me. For me a lot of the questions just didn't seem to feature where I am and the sorts of things I can and can't do". (P15, page 5, lines 189-191).

"Er ... well I suppose because I'm not in a great deal of pain and things don't change that much with treatment… I don't expect them to at my age but I want to keep going. I don't think they'll pick up much change at all actually". (P6, page 1, lines 27-29).

"Well they seem to me to miss things. I mean when I go for treatment... and you… well you know what is wrong with me when you look at me and know what needs

doing so I would suggest that you know through your expertise and you can't measure that"… (P9, page 4, lines 137-140).

1.3 Timeliness for completion

Experience of completing previous questionnaires made some patients cautious about the amount of time they would be required to spend completing the PROMs and other questions contained within the app. Others were happy to spend longer amounts of time.

"I wouldn't want to spend a lot of time on it...erm five minutes would be ok". P7, page 1, line 18).

"Five or 10 minutes… no hang on 2 minutes … as quickly as possible".(P5, page 1, line 41).

"15 to 20 minutes". (P11, page 1, line 26).

There weren't any notable associations between participant characteristics and the time suggested as acceptable for PROM completion.

1.4 Variety of formats (words and numbers)

A natural affiliation for either words or numerical scales was noted by some patients. The use of numerical scales in commercial feedback mechanisms was noted by some patients, while others found it hard to quantify their symptoms and found descriptive statements more helpful and relevant. Strong preferences were voiced for numerical scales in younger age groups who were more familiar with such scales, whereas older age groups tended to prefer descriptions in text only.

"I find it hard deciding on something that is 0 to 10 or 1 to 10… well I think". (P9, page 4, lines 127-128).

"It may be that literally every questionnaire I do I fill in is using a sliding scale. I've just had some sent to me from Sky and Virgin surveying me about the sort of service I receive and they have both used a sliding scale and it's really quick and easy to fill in. And then you just had your scale between 1 and 10 and then a comments box at the bottom and it's fine". (P15, page 6, lines 206-211).

1.5 Technological issues in completion modes

Familiarity with technology varied irrespective of age. Suspicion and aversion to technology was evident in some participants, while a willingness to "have a go" was evident among others.

"It would have to be paper as I don't really use the Internet... I wouldn't know how to turn the computer on… I leave all of that to my husband the Internet is his toy. I just don't touch it". (P9, page 1, lines 26-28).

"A phone app, well now then… Erm…well yes… I would have a go at that". (P7, page 2, lines 46-47).

"I don't really know what an app is to be honest with you" (P7, page 3, line 89).

Suspicion about technology was more notable in participants who did not have a computer and did not use the Internet (P6 and P10). Participants who only accessed the Internet or who had Smartphones but did not use apps noted that they would try to use them if necessary (P7 and P9).

Theme 2. Data protection

Some participants were very conscious of the use and potential for miss-use of their data. Strong views were expressed concerning commercial exploitation of data. Some of the interviews took place around the time that concerns about the "Care data" initiative were being debated in the media (NHS England, 2014). Strong views were expressed by one participant concerning potential sale of data to pharmaceutical companies (P7) but others stated their trust in their osteopath that they would not ask patients to contribute data which would be used other than for professional development (P2).

2.1 Choice of participation

Concern was expressed that some patients would be excluded if the modes of completion for PROMs were solely electronic. This was an issue of both access to technology and familiarity of use.

"You know it's important to be inclusive still even though recognising technological developments". (P5, page 3, line 112)

"I'm easy, I would do either but these things are moving much more towards the erm sort of app-based questionnaires and surveys and stuff so I suppose that is the way it's going".(P1, page 2, lines 45-47)

"I have no preference… I think each have their drawback... er certainly men of my generation aren't that familiar with mobile apps but I might have a go. I wouldn't have any problem with online but I would have to remember to log in and do it. Pen and paper I need to find the pen and paper so there's hindrances in all of them…" (P5, page 2, lines 45-48)

2.2 Data protection

Patients were reassured once they knew that no personal or other identifiable data would be collected as part of the data gathering process.

"… as long as they don't want your personal data … as long as you are just a patient a 35 year old man with a bad back… I don't have a problem with that. I wouldn't necessarily have a problem even if my address was on there I don't think but I could see how that could be a problem for some people". (P5, page 2, lines 65-68)

"Erm ... well as long as it is collected relating to the Data Protection Act and they don't disclose it without my consent to disclose it, but if they were to disclose it somewhere they must ask for my consent before they disclose it. I disclose it based upon the treatment I'm receiving I wouldn't like it to be shared all over the place. You know when I tick the box that I'm depressed I wouldn't like them to disclose it to any other organisations without my consent because sometimes people don't understand what it's like and what sort of pain you're going through and they might have negative views on me". (P2, page 2, lines 44-51)

"It depends what the purpose would be… erm… and what this research group was going to use this research for…I don't suppose in principle I would object to that if I thought it was going to help the next person in their treatment". (P7, page 2, lines 51-53).

2.3 Use of data

There were some strong views expressed about potential commercial exploitation of patient data. However there was also trust that osteopathic practices and a university-based study would not permit such data sharing.

"Yes, I think so. Erm also and I think that data protection would need to be stressed … that information for example wasn't going to be sold on… I'm not saying it would be... but for example say there was a pharmaceutical company that was selling I don't know what… but they might be very interested to know that you would buy one of

those if it would be helpful and I would want it to be very clear that my information wasn't going to be sold on to other people who would contact me and say 'ey up do you want to buy one of these things'" . (P7, page 2, lines 56-62).

"So I think we have a say in what we want to do, so with me giving consent I know that you're using my details and I'm telling you things I'm happy to disclose but if I'm not I don't want anyone taking my details and then using that". (P2, page 2, lines 77-79).

2.4 Confidentiality of technology

Some concern was expressed about where the PROMs would have to be completed *i.e.* in practices or at home.

"No, I'd prefer to do it on my phone. While I'm travelling back, I'd prefer to make use of that time, and to do it on my phone using an app or even online I'd still do it over the phone". (P2, page 1, lines 38-39).

"Erm… well... I think there would be some pressure sitting in the practice... erm I mean it could be emailed... it could be emailed to me in advance of the appointment and you could sit and ponder and really think about the answers rather than just give an immediate response". (P7, page 1, lines 21-24).

"I'd be quite happy to fill it in in the practice. It wouldn't stop me saying something negative if that's what I thought". (P3, page 1, lines 21-22).

Theme 3. Purpose of PROMs

Some patients expressed the view that there had to be a clear statement about why data were being collected to ensure their participation. It was recognised that data collection was a commonplace event in commercial organisations and retail outlets but this was less prevalent within healthcare. Others expressed the view that although they had completed questionnaires in the past they failed to see how that had manifest in changes or improvement in treatment/overall care which was their main concern.

3.1 Clear statement of purpose for data collection

Participants felt there would be better completion rates if it was clear that data were being collected for specified purposes, and this information made explicit by notices displayed in practices.

"I guess it's the same with all sorts of data collection that the thing is that with all of the data it should be clear what exactly is going to be done with it, and why do we want to know this otherwise why are we bothering". (P1, page 2, lines 65-68).

"If there's a clear reason and some sort of benefit I can see. I mean not like supermarkets collecting data from my loyalty card that sort of thing. I think that sort of thing is a bit intrusive". (P4, page 3, lines 75-77).

"Well it should be useful but I don't know whether it is as nothing seems to change does it?" P6, page 2, lines 63-64).

3.2 Feedback within the consultation

Some participants felt that the PROM data would be more useful if it was shared within the consultation allowing the osteopath to be aware of progress or lack of it and could amend treatment strategies accordingly.

"So... I'm sorry but.. I don't see the point if the practitioner isn't going to see either form". (P7, page 1, lines 33-34).

"I think anything that you try and do which allows you to see... and you know kind of like a weight loss chart … if you can actually see it written down... you know that's where you were and that's where you are now and you have actually moved on and progressed…". (P1, page 1, lines 14-17).

3.3 Measuring treatment effectiveness

This was felt to be important in identifying whether treatment was being helpful or whether different treatment or referral was indicated. In addition, such data was seen to be important to engage other professionals and funders in the patients' perceived merits of osteopathic treatment.

"Things do seem to be finally moving in what I would call a more scientific approach. I imagine it is difficult to get osteopaths who are all working privately to contribute to this type of system and move this type of thing forward, but in the long term it would do the osteopathic profession quite a lot of good if it had a bit more scientific data and papers published and all of the rest of it". (P4, page 8, lines 312-316).

"I think it's really helpful to see something written down that actually shows you. And I suppose that these days because it's been a positive... there's been an improvement that's where I am now but I suppose also it's helpful if there's little or no improvement then I guess you have to start thinking about what else you're going to do. What can you do: what other options have you got".(P1, page 1, lines 17-22).

"It's the only way they're going to find out whether the treatments they're giving you are actually working… erm ... and it takes a lot of info to actually come up with some figures at the end of the day". (P13, page 1, lines 21-24).

3.4 Consideration/reflection on other issues

Views were expressed that the content of the PROMs made participants reflect on different aspects of their symptoms. This was viewed as a positive activity; some participants expressed the view that their symptoms and limitations were not as bad as the descriptions in some of the PROMs. Other participants stated that the content of the PROMs raised issues about items they could discuss with their osteopath.

"Having to hold on to something to get out of chairs that's all sort of stuff that is alien to me at the moment certainly... I feel better for knowing that". (P1, page 5, lines 184-186)

"I think it highlighted err something that you wouldn't normally associate with it.. and I would go... 'oh yes, that does happen to me' whereas I tend to think they are just normal things to do". (P13, page 5, lines 177-179).

"I mean here with the one about turning over in bed, that's one of the questions I might have answered yes to but it's not that I find it difficult but I realise I make myself aware that I need to do it carefully or it catches if I don't". (P15, page 5, lines 180-184).

Theme 4. Motivation

Participants expressed a range of views and about motivation for completing PROMs data, and the types of information they would like to see disseminated from that information. It was recognised that contributors of data could have positive or negative reasons for submitting data, and they may not be representative of the general population of patients attending osteopaths. For other participants there was clearly strong motivation to highlight their views about the benefits of osteopathy thereby potentially increasing access to treatment through state-funded healthcare provision.

4.1 Information about practices (*e.g.* types of problems being treated)

Some participants expressed the view that information about the conditions osteopaths manage could be more explicitly displayed in practices.

"It's more to do with this osteopath sees most patients with this condition and they take an average of so many treatments, and they use other treatments in their management. Or they use these sorts of approaches". (P1, page 3, lines 83-86).

"If there is information about practices, perhaps if they had specialities, and this sort of thing. I think some do have particular specialities. I know my particular practice is specialised with horses so um.. I think specialities particularly could be useful. I know some specialise in cranial osteopathy for babies I think that's something that should be known". (P14, page 4, lines 133-138).

"What I'd like to see is what they are offering… yeah what they are offering and as a percentage how many patients they have that would be quite attractive as I would know that that clinic is doing a lot of back pain". (P2, page 3, lines 111-113).

4.2 Widening patient access/choice

Strong views were expressed by some patients about the benefits they had received with osteopathic treatment compared to mainstream/allopathic treatment. Many felt that participation in data collection could contribute to widening access to osteopathic care through increased public funding.

"Because it's still difficult with some of the GPs isn't it. But if NCOR can start to chart how patients are doing then maybe when you get this data together from patients and they see what we say it might make them change their minds. If it was fairly concise and didn't have to take too much reading of it, concise like a bar chart sort of thing, that might be helpful". (P9, page 5, lines 163-171).

"I just really want to help osteopathy and I just wish it was part of free healthcare". (P14, page 7, lines 263-264).

4.3 Information for other healthcare professionals

Some participants expressed the view that there was considerable room for education of healthcare professionals about osteopathy.

"Because if you go to the doctor they don't recommend them do they most of the time. They don't! I mean I don't go very often but if I ask the doctor he'll say, 'No, I don't believe in it' or whatever and they are actually the ones who need educating not us". (P9, page 5, lines 157-160).

Theme 5. Dissemination and feedback of information

Participants were asked during their interviews about how they would like findings from practice-based data collection to be fed-back to them. Issues about excessive amounts of information, laden with jargon, and ultimately largely unintelligible were raised. The over-riding sentiments expressed were that dissemination should be succinct, expressed in lay language, and available through a variety of media.

5.1 Information overload

The view was expressed that too much information was available for patients both in written form, and available through the Internet. The volume and content of the data could feel overwhelming and had the effect of disengaging the patient from receiving information. Any data from the study should be succinct.

"Erm… gosh… I really haven't got a strong view about that but the simpler the better. Erm if it could be graphical ... if it would have some images… not too many medical words to describe what the injuries have been ... but you're talking to someone who is just Joe Bloggs on the street and not another practitioner. I suppose it's a bit like a bedside manner but as a written bedside manner... do you understand what I mean". (P7, page 3, lines 83-87).

"Say one half piece of paper with the information and a little graph or chart in the middle, a diagram and the contact email address of the person who did it". (P2, page 3, lines 103-105)

"I have asked whether my condition is very common and he has discussed in general conversation the type of questions I have asked so I suppose to see it on a poster that would be quite helpful ... and you know to find out that the majority of people or X amount of people in the practice suffer from whaah… whatever it might be. I think that would be quite helpful". (P1, page 3, lines 89-93).

"Erm I think you don't want long period of text, you want to put it in a short succinct way erm and a graph and some recent statistics would be fine". (P5, page 3, lines 87-88).

"I should think most people just want to get their back fixed and then off they go". (P11, page 3, line 79).

There were no specific associations between participants and their views on information overload. A combination of visual and written text was regarded as the best option in most age groups (P1, P2, P5). For other patients who had received osteopathic treatment over a long period and regarded it extremely favourably, they felt that data for patients was of less value than data for other healthcare professionals potentially to increase access to funded osteopathic care (P11)

5.2 Variety of dissemination formats

Participants felt that information about the PROM project should be disseminated in a variety of formats accessible to all patients but should be brief.

"What I would say to you is that if they had a poster in the practice itself it attracts all sorts of people that they are seeing who are going in and out. They could send it out as an email to all of the patients and some might look at it and some might not but that's also a way of communicating this so I would read it yeah". (P2, page 3, lines 88-92)

"Yeah I think again it would be a leaflet that just sort of headlines the conclusions. I'm not particularly interested in knowing how the data was used and all of the rest of it. I guess that's of more interest to the osteopathic profession. To me the only thing of interest would be the outcome and how it related to me". (P4, page 3, lines 98-101).

"I think really on a website". (P15, page 4, line 135).

Once again there were no particular associations between dissemination format and participant characteristics. Younger people (31-40) and (41-50) who were working and highly IT literate regarded multiple forms of access as helpful for different reasons including convenience and increasing exposure of osteopathic care (P15 and P2 respectively).

5.3 Clinical versus research purposes

Many participants expressed the view that the PROMs could be very useful as a clinical tool to aid discussion and measure progress in the resolution of their presenting complaint(s). Some were less convinced about the value of independent data collection alone, and did not appreciate the potential for bias in of data if data were shared with clinicians also. The use of PROMs as clinical monitoring tools is being established increasingly especially among manual therapists, and it may be that the message concerning the need to collect data about the effect of treatment needs to be more clearly expressed.

"People should respond honestly I think whoever is going to see it whether that's the practitioner or not because if treatment hasn't worked you need to go back or find somewhere else but they need to know that". (P7, page 2, lines 38-40).

"Well I guess if you weren't that happy with your osteopath you wouldn't be there anyway… but you know unless someone was breathing down your neck or watching you … I mean that wouldn't be appropriate…but if you were just given this to do and the time you needed to do it so you could hand it back in as you were booking your next appointment then as long as someone's not standing over you then no, that's fine". (P1, page 2, lines 52-57).

None of the participants stated that they would be unhappy to reply in a forthright manner concerning the effect of their treatment on their symptoms. Honesty was seen as being more valuable for both patient and clinician concerning treatment progress or its lack, or about any practice issues which were regarded as unsatisfactory. These comments were made by patients of older age groups (P7 and P11). Another participant in a senior management role (P1) also noted that a straightforward approach was of more value to the professional encounter.

2.3.3 Main constructs and summary

The main findings can be summarised more clearly and succinctly by grouping them into three main constructs:

1. Attitudes to data collection;
2. The experience of osteopathic care;
3. Concerns about data capture and purpose.

Each construct will be discussed in turn.

Attitudes to data collection

Participants described a range of attitudes to data collection and the potential value they felt it could have for osteopathy as a discipline. Some participants commented on the development in the profession during many years of experience of care, and the need for a maturing profession to become more organised in how it examined the effects of care and its delivery. Comments were made about the value of being able to describe day-to-day practice to a range of audiences: sometimes this included existing patients who were not always entirely familiar with all of the different symptoms or symptom areas treated by osteopaths. A different aspect to this attitude was the belief that collecting a body of data about osteopathic care would help to educate clinicians who were either treatment-naïve, ambivalent, or hostile to the idea of osteopathic management. Conversely, the possibility that healthcare professionals who were receptive to osteopathy would have greater evidence with which to advocate funding for care was not discussed by any participants.

The manner in which data were collected was highlighted by participants. While some individuals who were experienced and confident in using technology were quite open to the idea of electronic data capture. For other participants who had less access to IT and were less familiar with IT innovation in the form of apps this was seen as an area where they may require support from their partner or children but nonetheless were willing to try to use it.

A smaller number expressed apprehension and antipathy towards any type of IT device, and exhibited no desire to change that when the issue of providing support was suggested. In a small number of cases older participants did not own a computer or have access to the Internet by any other means.

The experience of osteopathic care

Interviewees highlighted that their experience of osteopathic care involved more than just being recipients of different techniques. Different dimensions of care were important including the feeling of being listened to, empathy with the clinician, and the opportunity to ask questions about their symptoms. The role of

education and advice and its importance to patients has been explored in other surveys. This supports clinical guidelines where the role of self-management and patient empowerment are advocated. Participants expressed their support for data collection since they felt the treatment encounter was valuable due to these separate elements. Collection of data was seen as important to facilitate widening the access to such care, and holistic patient management.

The value of collected data to provide objective and systematic findings when communicating with other healthcare professionals, and commissioners of services was remarked upon by many participants, particularly when there had been a journey through other forms of healthcare management prior to osteopathy. Other participants noted that the organisation of the profession for data collection was long overdue and this was a surprising omission from the clinical encounter. For some participants, their experience of osteopathic care did not rely solely on the management of musculoskeletal conditions. This was one of the reasons why participants' views on a small selection of PROMs were sought. Patients experiencing a range of disorders, both musculoskeletal and non-musculoskeletal, will nonetheless experience some common symptoms *e.g.* anxiety, disruption to the working and social world, and it was important that these wider considerations were captured in any PROMs used in clinical practice. There were, however, some caveats accompanying the enthusiasm for potential use of patient data and these are described in the next section.

Concerns about data capture and purpose

While patients recognised the value of collecting data, especially if that might facilitate access to osteopathic care funded via the NHS, allowing ineffective treatment to be identified, and supporting early referral where appropriate, there were some concerns about the disruption to the clinical encounter. One sentiment that was expressed was that the data collection would become the focus of the clinical encounter instead of the patient's wellbeing and meeting their expectations of care. Other views were expressed concerning how well PROMs could capture symptoms which had inherent variability from day-to-day but sometimes within the day.

For other participants it was important that any initiative should be translated into better patient care. The tension between feeding back data which could be used during consultations, and the importance of gathering independent data to develop the profession in a range of ways was not raised.

2.3.4 Evolved models and their meaning

Participants expressed a range of views based on their experience of attending one or several osteopathic practices for treatment over a number of years. These experiences, and those with other healthcare providers, were both satisfactory and unsatisfactory. However, based on previous experiences of data collection in other arenas, and with national changes in sharing data highlighted in the media this was regarded with some circumspection concerning the use of the data. Coupled with this were some concerns about the actual benefit such data collection activities would produce on the delivery of care.

The role of trust in the clinical encounter and the relationship with the osteopath was raised by some participants. In many cases the relationship with their osteopath had been of many years standing even though there were periods of no treatment while patients were symptom free. Any involvement in data collection was regarded within the context of that trusting relationship, where the osteopath could be regarded as acting as agent for the data collection. An assumption was that if they were involved it would be fine, and the misuse of data would less of an issue. However, in view of the way data are collected on a regular basis by a variety of different organisations, participants felt the purpose of the data collection and its potential use should be absolutely explicit to remove any doubts. The choice of participation for some participants was largely governed by their access to and familiarity with IT. Other participants expressed the sentiment that they would have been happy to participate had paper data collection been involved.

Younger participants who were familiar, frequent, and competent users of IT (digital natives) recognised some barriers to completing questionnaires. While

some expressed the view that they would be happy to complete a PROM on phones while on their journey home from treatment while on public transport, others who were fitting in appointments between work and family commitments recognised that they might put the task aside for later and it could easily be forgotten. A smaller number expressed the view that all they wanted was to "get their back fixed" and lacked motivation for further activities. Older patients who had attended osteopathic practices for some years after receiving various other forms of care that were unsuccessful were the greatest advocates for routine data collection, with some expressing the view that the profession was perhaps lagging behind other similar professional groups. They expressed frustration sometimes that their experience of medical professionals had identified a lack of knowledge about osteopathic practice, and the lack of potential benefit that this intervention could offer based on their own experience.

Completion of an electronic system that involved providing an email address raised a concern from some participants that they would be bombarded with emails in the future and this was the last thing they wanted. This underpinned their reluctance as they did not want continual pestering emails, irrespective of the potential value of the initiative. Once they were reassured that this would not be the case they were more receptive to the use of electronic communication as part of the data collection process. Other views were expressed, notably among older patients who had undergone extensive periods of treatment that sometimes the pain can be so debilitating that they suspected they might not feel like completing a questionnaire. The effort involved in dealing with the activities of daily life used up their capacity to do any form of activity and they might decline to participate in data collection however reluctantly. These relationships are summarised in Model One (Figure 2.1).

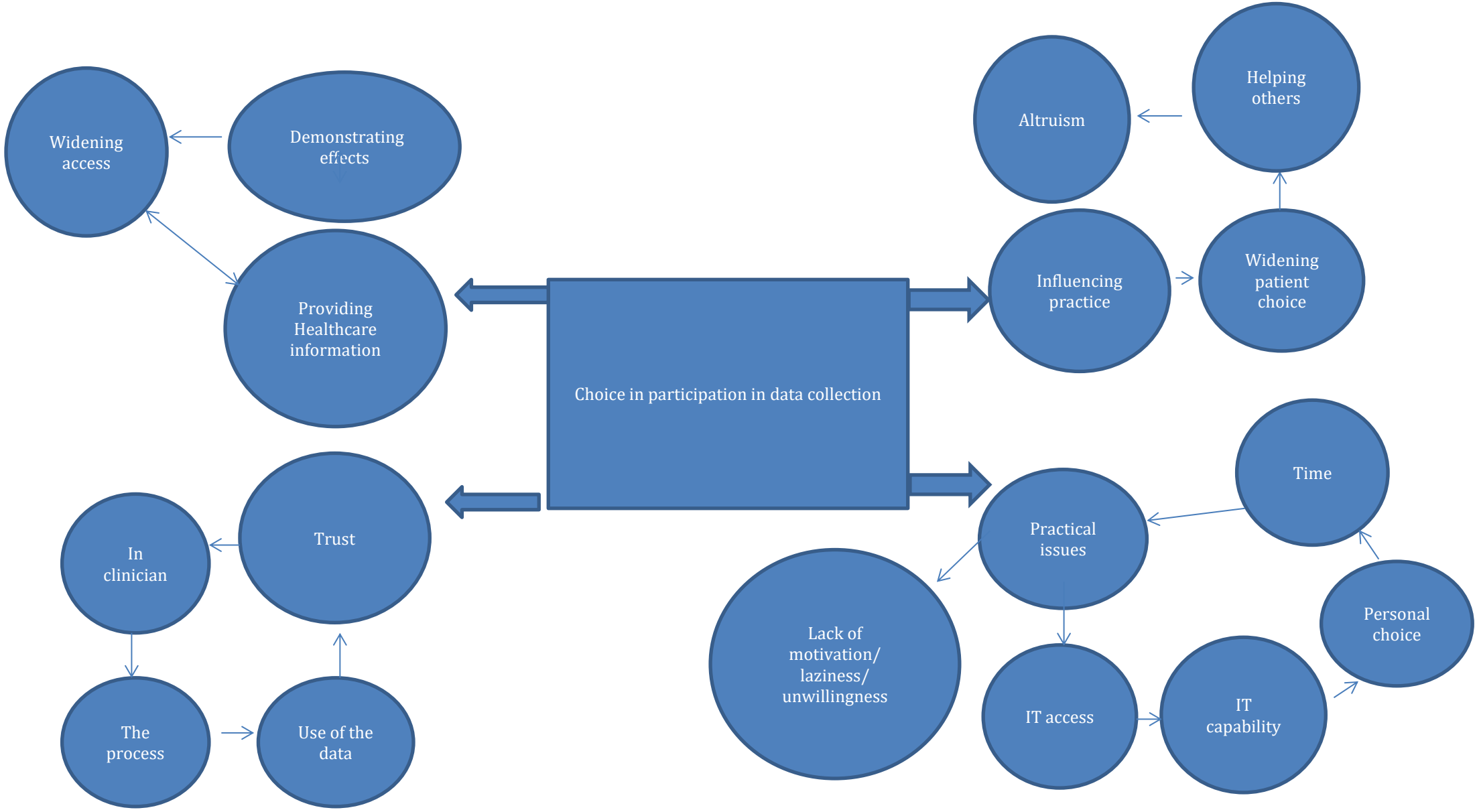


Figure 2.1 Model one: Examining the factors that influence patient participation in practice-based data collection.

While participants contributed many views on the benefits, barriers, and facilitators to practice-based data collection, the actual value of the data was discussed also. A second model (Figure 2.2) was derived from the qualitative data which addressed the usefulness of data collection to individual participants, practices, and the profession as a whole.

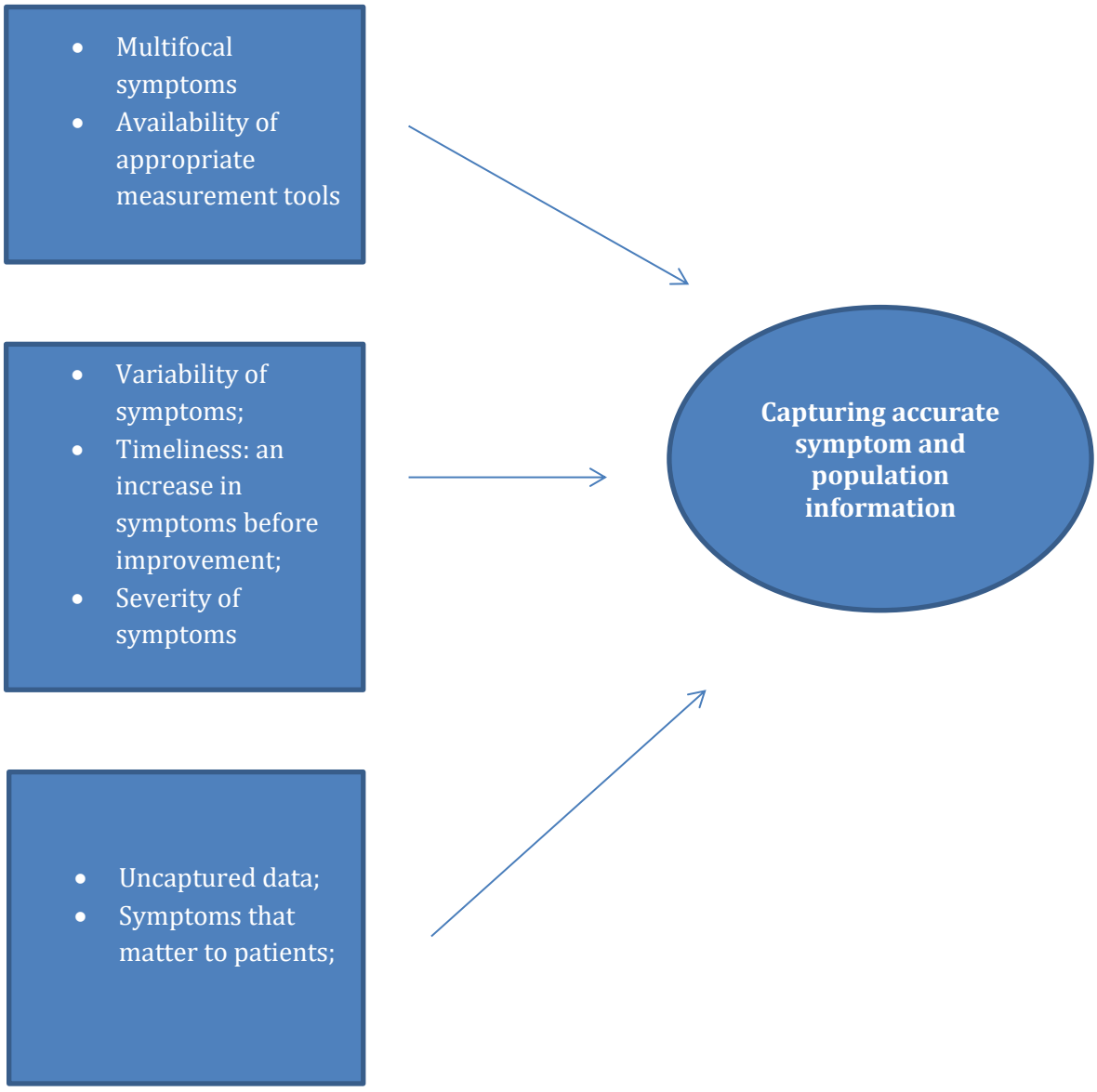


Figure 2.2 Model Two: Issues around the usefulness of data collection

While Model 1 describes the factors that influence patients in participation in practice-based data collection, Model 2 addresses issues around the usefulness of data collection. A range of issues was raised by participants who were asked about

their views on the value of data collection generally, and their views on three specific PROMs. Patients attend osteopathic practices for a variety of reasons, and osteopaths practice in a range of different "styles" encompassing a distinctly musculoskeletal approach whereas others employ techniques which work on the "involuntary mechanism", and viscera, for example, to attempt to address non-musculoskeletal symptoms and conditions. Some of the participants in the interviews (*n*=4) described their symptoms as being non-musculoskeletal and expressed their perceived value to this. They noted that it would be impossible to capture data on all the variations of their symptoms but did not see that as a limiting factor in data collection.

When examining what some of the PROMs were collecting, other participants expressed the importance of using a suitable tool. There were differences in views of the format for the PROMs (numerical or text) but for some participants they felt that there were insufficient options in the options listed (the Oswestry Disability Index), or that the statements used in the PROMs were targeted at populations with much more severe levels of disability (the Roland Morris Disability Questionnaire). The Bournemouth Questionnaire was viewed most favourably, and participants noted that the statements within this PROM touched on issues which were important. However, some participants noted that none of the PROMs had been able to capture the sheer frustration associated with musculoskeletal symptoms, and the disruptive impact it can have on family life, especially when it came to having to ask for help.

For other participants, they expressed views about the variability of symptoms. While recognising that any questionnaire must capture data at a point in time, they felt that none of the tools they examined, or in some cases had previously encountered, could reflect this diurnal symptom variability.

2.4 Discussion

This qualitative study has explored a range of different issues associated with data

capture, and asking patients to complete patient reported outcome measures within a private practice setting. The main findings are outlined in the next section.

2.4.1 Main findings

The results of this study show that the participants interviewed expressed their broad support for the idea of practice-based data collection. There were differences of opinion concerning the most useful PROMs depending on whether participants had a particular affiliation with words or numbers, but bpad agreement on the relevance of those selected and the importance of relevevant data being collected to facilitate the data collection process. The desire to use an electyronic data capture system was not universal but pateints indicated they would be willing to complete it in a suitabke setting where help could be sought if required. Patients noted that use of the PROMs provided some reassurance about the nature of their symptoms, but also had their awareness raised about areas for further discussion of symptom management with their osteopath.

The practicalities of data collection were highlighted by some patients focussing on the need to translate good intentions into action, and making the process as quick and user-friendly as possible. In a small number of participants, the use of their data was discussed in greater depth especially for use in commercial use and being fed back to osteopaths. At the time of the interviews the potential for sharing data collected in primary care was topical in the media and some patients were concerned about what they regarded as the potential miss-use of their data (NHS England, 2014). Patients recognised the potential opportunities for pan-professional data collection (*e.g.* education of other healthcare professionals) but were less interested in accessing local practice-related data. If any feedback was available, for example on a practice website, patients suggested it should avoid technical jargon, too much text, and perhaps have one or two diagrams (a bar chart was suggested). There is some overlap and some contrast with the work by Hildon *et al.,* 2012. Participants in the Hildon qualitative study preferred tables, and thought that bar charts were confusing. There was agreement on brevity "less is more", and the use of contextual information to allow patients to make sense of the data presented (Hildon *et al.,* 2012).

There are other areas of research where this current qualitative data can be contextualised. This is discussed in section 2.4.2.

2.4.2 Previous research

It is a notable contradiction in healthcare is that continual change is a constant feature. Less constant in healthcare has been the changing role of the patient. Thompson (2007) described the shift in the balance of patient involvement in various stages from a parentalist model where patient involvement was limited to receiving information, and giving consent where necessary to the model of informed decision-making where the patient is the final arbiter of decisions where treatment or intervention options are offered. The dissonance within these models is clear in the qualitative contributions from the participants in this study. While some patients are very keen to have their voices heard and view PROMs as a way of facilitating the process, others prefer to maintain a parentalist relationship. The role of trust has been described by many contributors and this may underpin the retention of this parentalist model in many of the patients, especially among the older age groups.

Over the past decade, concern in publicly-funded healthcare has shifted from measuring patient satisfaction to measuring experience of healthcare also. Many satisfaction measures are based on expectations that may be very low, and essentially insensitive in detecting shortcomings (Black and Jenkinson, 2009). Gathering patients' experience through feedback offers an insight to the effect of care, and its humanity (Black and Jenkinson, 2009). However, the relation between patient experience and outcome is unclear; we do not know whether a poor experience increases the likelihood of reporting poor outcome or vice versa (Black and Jenkinson, 2009). A recent survey in 2015 (n=1,566) commissioned by the General Osteopathic Council, members of the public and existing osteopathic patients were specifically asked about their willingness to provide feedback on their experience of care, and their preferred mode to accomplish this. The findings are shown in Table 2.5.

Table 2.5 Public and osteopathic patients' preferences for providing feedback about osteopathic care.

Method of feedback provision	General public	Existing osteopathic patients
Online survey completed at home	44%	35%
Face-to-face with the osteopath	25%	47%
An anonymous form at the practice	18%	10%
Online survey completed at the practice	9%	6%

Source: General Osteopathic Council. Public Perceptions Study, 2015. http://www.osteopathy.org.uk/news-and-resources/document-library/research-and-surveys/public-perceptions-study-full-report/

This collection of patient feedback is largely absent day-to-day in both publicly-funded healthcare and osteopathic practice. However, it is becoming increasingly challenged by funders of care *e.g.* health insurers, and acts as a barrier to accessing new opportunities for care delivery *e.g.* the Clinical Commissioning Groups. The greatest merit for the data for some participants was in educating other healthcare professionals. Research by the GOsC suggests awareness about osteopathy continues to grow. In surveys conducted in 2001 (n=413) and 2006 (n=1003), 13% and 14% of patients respectively had learned about osteopathy from either their GP or practice nurse (GOsC, 2001; GOsC, 2006). In 2009, the British Osteopathic Association (now known as the Institute of Osteopathy) produced a report as a result of a Freedom of Information Request (Freedom of Information Act, 2000). They requested information from every Primary Care Trust (PCT) in England during summer 2009 about their use of osteopaths within their local NHS settings. The responses indicated that while 16% of PCTs responded that they allow GPs to refer patients to Osteopaths on the NHS, only 15% of PCTs gave evidence of funding

any patients for osteopathy in 2008-2009. In areas where there is favourable awareness of osteopathy and provision the number of patients referred has increased year on year, indicating GPs and patients are pleased with the results. The numbers of treatments funded varied widely among PCTs with one patient funded in Lincolnshire Teaching PCT in the year 2008/9, to 1970 patients in South West Essex PCT in the same year (BOA, 2009).

Patients and new technology

The importance of involving patients in new initiatives is clear. Kaplan and Maxwell identified that computer information systems can significantly improve patient care, management, practice administration, research, and medical education, but Dowling estimates that 45% of computer-based medical systems fail due to user resistance (Dowling, 1980; Kaplan and Maxwell, 2005). Many evaluations of new electronic medical innovations focus on costs and benefits, completeness, error rates, retrievability, usage rates, user satisfaction, and clinician behaviour changes (Kling and Scacchi, 1982; Kaplan , 2001a; Kaplan , 2001b). However, since the work of Dowling, and Kaplan and Maxwell there has been a greater embedding of technology within our day-to-day existence often in quite subtle ways. The value of using new technology in data collection in a NHS setting has been studied recently by Malhotra *et al*., 2016, and Jenkins *et al*., 2016. Malhotra *et al.* found that technology was implemented successfully into their elective orthopaedic setting with a completion rate of 85.9%. In contrast Jenkins *et al.* reported that only 72% of their patients attending an orthopaedic clinic had internet access concluding that e-data collection had the potential to introduce bias due to exclusion through lack of access based on age and social deprivation (Burton, 2013). Both of these conclusions are worthy of comment. In my qualitative study lack of access to IT in older patients was sometimes due to lack of interest or preference rather than income. Mobile only households are concentrated among those with lowest income (47%), but perhaps this makes the case for the choice of surveys as web and mobile apps (OFCOM, 2009). Older patients in this qualitative study stated that they would seek help from family members as required. Perhaps a greater challenge in terms of access is the potential language restrictions faced within an increasingly diverse society where individuals may be able to speak

English or an available translated language but have literacy difficulties (ONS, 2012).

While society now recognises the importance of obtaining the views of users in developing services like electronic data capture in healthcare (Wensing, 2000; Wensing and Elwyn, 2003), one issue with assessing preferences about what to include in those systems is that patients' decisions about what is important in healthcare are often based on their individual experiences rather than reflecting on the concerns of patients in general (Uhlmann *et al*., 1984, Deber, 1994; Mullen, 1999; Wensing and Elwyn, 2003; Staniszewska *et al.,* 2012).

Challenges for clinicians

The views of patients expressed in this study highlight challenges for clinicians. Many patients wish to take part in the decision process when their healthcare is involved (Guadagnoli and Ward, 1998). While patients agree on the principle of collecting data and can see the widespread value of this approach particularly in the role of advocacy for the profession, there is also the tension of being asked to fill in another form. In a world where patient and consumer opinion is being sought increasingly, a fatigue factor can start to occur which then undermines the whole process: this sentiment was noted by patients requesting a brief questionnaire, which was quick to complete. Tensions also arise as described by some interviewees about the purpose of patient data when it does not become an implicit part of the care process. While PROMs as monitoring tools fulfil a valuable role, they cannot easily be both monitoring tool and unbiased measure of effect of treatment at the same time.

Dealing with practical issues

The introduction of routine PROM data collection produces a range of challenges. These can include issues of implementation, patient involvement in selecting instruments, timeliness, and interpretation of data. Clinicians may lack the skills to implement and use specific instruments or have negative attitudes about specific approaches. While this can be addressed by incentivisation frameworks in some organisations their effectiveness is yet to be fully investigated (General Practitioners' Committee, 2002). The correct use of instruments, and

consideration of recruitment bias may also pose an issue for clinicians. Methodological issues with respect to adequate data collection in terms of sample size, and appropriate case mix must also be considered when PROMs are used in research settings as compared with day-to-day clinical practice (Jenkinson and McGee, 1998). Although guidance from systematic reviews can inform clinicians about the most appropriate PROMs, that is not always sufficient to address concerns that PROMs are neglecting what are important to patients (McDowell, 2006).

Desirability versus realism

When planning new innovations it is important to consider the benefits but also unintended consequences that could arise. These could include:

- Unrealistic patient expectations of what healthcare can deliver;
- Subsequent dissatisfaction if those unrealistic expectations are not met;
- Defensive behaviour of care providers resulting in higher numbers of unnecessary clinical procedures or examinations;
- Undermining of professional morale;
- Increased costs associated with the above issues (Wensing and Elwyn, 2003)

The prospect of limiting practice has been identified in the literature as another unintended consequence. Chew-Graham *et al.* investigated the role of outcome in routine review consultations in primary care (Chew-Graham *et al.,* 2013). They identified that "*routine review consultations in primary care in patients with chronic conditions focus on the biomedical agenda set by the Quality and Outcomes Framework (QOF) where the practitioner is the expert, and the patient agenda unheard. Review consultations shape patients' expectations of future care and socialize patients into becoming passive subjects of 'surveillance'. Patient needs outside the narrow protocol of the review are made invisible by the process of review except in extreme cases such as anticipating death and bereavement*". The use of PROMs in this instance could highlight the needs of the patients and identifying factors about their symptoms and management that are important to them.

Clinicians' and health commissioners' views concerning the use of PROMs have been explored in a qualitative study (Hunter, 2014).

The data from this study has revealed a range of interesting challenges for the development of an electronic data capture system. To feel reassured about the robustness of the findings, it is important to discuss the validity or credibility and reliability or trustworthiness of the qualitative data also. This will be revisited in the next section.

2.4.3 Strengths and Limitations

Despite the revised focus on patient-centred care, and the changing role of the patient documented in the literature, there is a notable lack of qualitative research involving patients and the use of PROMs. Patients have been included in the development of many PROMs, but views concerning where and when they should be used, completion times, and the availability of different formats are notably absent. One key feature of this PhD was for the creation of a data collection system designed for patients to use. To create a system for them without their input would have been unimaginable and foolhardy. There were mixed views on the use of feedback within the consultation. Some participants expressed the view that this was important for them. This is a feature of the data capture system that will need revisiting in the future.

Inevitably, there will be limitations to any research. The sample obtained for the interviews contained greater numbers of older and retired patients than any other group. Although the sample reflects the population identified in previous data collection work, making the data collection system as accessible as possible to all potential patients benefits from input from a wide cross-section of individuals. This was shown in the sampling characteristics of participants in Table 2.3 (Burton, 1981; Pringle and Tyreman, 1993; Hinkley and Drysdale, 1995; McIlwraith, 2003; Fawkes *et al.,* 2012).

The decision to conduct the interviews by telephone was a practical issue to achieve completion of this aspect of the study within a defined academic programme. While

enhancing the geographic spread of the participants was achieved, it meant that nuances in facial expression and body language that can add to the richness of qualitative data were absent (Garbett and McCormack, 2001; McCoyd and Kerson, 2006). The importance of face-to-face as opposed to telephone interviews is much debated in the literature (Novick, 2008; Sturges and Hanrahan, 2004). There is a lack of comparison studies, and there are clear views on the value of varying the approaches for different types of data, subject areas, and populations (Groves, 1990). Among the small number of head-to head studies, very little difference in the data has been revealed (Tausig and Freeman, 1988; Sturges and Hanrahan, 2004). The interview approach must be employed with consideration to widening access for both participants and researchers, comfort of the participant, researcher safety, and cost (Aday, 1996; Chapple, 1999; Carr and Worth, 2001; Bernard, 2002; Novick, 2008).

The interviewing of osteopathic patients by an osteopath can produce information that they may feel will be well understood. However, an interviewer who was not an osteopath may have picked up on more issues where patients felt osteopathic practice could be improved. There are tensions in an osteopath developing a system for osteopaths, and analysing the qualitative data that is an implicit part of that process. Considerable self-awareness has had to be present to avoid looking for answers that are desired rather than hearing what patients are actually saying. The need for an independent non-clinician to assess the qualitative data was important to produce some distance and criticality to the process when triangulating the patient qualitative data. In a similar manner, engaging an osteopath to develop a data capture system for osteopaths is helpful in considering feasibility, and identifying issues of importance to clinicians. However, there is the disadvantage that the final product could be too inward-facing and not constructed with enough awareness of the wider healthcare arena.

The concept of reliability (Ritchie and Lewis, 2003) or trustworthiness (Guba, 1981) has been identified in qualitative research. Guba proposed four criteria should be considered in pursuit of trustworthiness (Guba, 1981). In addition, Ritchie and Lewis describe the concept of validity in the Framework approach

Ritchie and Lewis (2003). For other researchers this has been described as credibility: it relates to whether the findings are believable (Guba, 1981; Lincoln and Guba, 1985; Bowen, 2009; Holloway and Wheeler, 2015). Methodological concerns which relate to trustworthiness and credibility within qualitative research include, for example, the sample coverage, the environment for capture of the data, the identification or labelling of the data, the manner of interpretation of the data, and the display of the data and how this remains true to the original data. These concepts are described in Table 2.6

The usefulness of the qualitative approach in this setting is that it gathers information on the meaning and context of the phenomena studied, the particular events and processes that make up those phenomena over time, in real life, and natural settings (Maxwell, 2006). When evaluating new IT systems such contextual issues include social, cultural, organisational and political concerns surrounding the IT system, the processes of IT system development, installation, the use (or lack of it), and how all of these different factors are conceptualised and perceived by the participants in the particular setting of interest (Kaplan and Shaw, 2004).

Table 2.6. Strategies used to enhance the trustworthiness of the study
(Guba, 1981; Lincoln and Guba, 1985; Holloway and Wheeler, 2015).

	Description	Strategies
Credibility	Confidence that the research has provided an accurate determination of the meaning behind the data thereby reflecting the patients' experiences of osteopathic care	**Immersion in the data.** Time was spent initially engaging with all of the patient data to become immersed in the richness of the finings; **Member checking**; A sample of participating patients were invited to read through the verbatim transcripts of their interviews to confirm its accuracy, and have the opportunity to request removal of any comments with which they were later uncomfortable. Peer debriefing: discussion of the findings with individuals (researchers and non-researchers not involved in the study to provide feedback on the coding and analysis of findings.
Confirmability and dependability	Whether the findings from the study provide a dependable and realistic presentation and interpretation of the views expressed by patients.	Audit trail. The importance of maintaining field notes, memos, and other recordings of the research process either in visual or written forms. Maintaining notes on the stages undertaken and decisions made during the analysis process whether singly or in collaboration with other researchers. The researcher was able to remain reflexive towards her involvement in the interview process, and the impact this

		could have on questioning and interpretation of the findings. Participants were reminded that their views were important and there were no right or wrong answers. Participants were asked to be as forthright as possible with their views on osteopathic practice and the potential use of PROMs in an osteopathic clinical setting. They were reassured that their data would be anonymised and at no time would they be identified in any form of publications of the findings of the research.
Transferability	The extent to which the ideas generated by this research may be applied to other settings and patient populations.	Writing accurate accounts based on the rich data provided by the patients, and providing information that could be applicable to a range of different research settings.

2.4.4 Future research

Patients identified many issues during this qualitative study which had clear implications for the app. Patients gave their views on preferred formats for questionnaires and the importance of relevant content. An acceptable time for completion for patients informed the need to be very specific on question content to avoid undue burden. Two PROMs were included at the pilot phase to meet the desires expressed by patients for outcome measures including statements for single responses (the RMDQ), and for those including a numerical scale (the BQ). The selected PROMs included a sufficiently wide range of response options to ensure relevance to patients and their symptoms (RMDQ and BQ). Patients did not feel that the ODI offered enough options to evaluate their responses.

A clear statement on the limitations of use of the data was provided within the patient information prior to completion of the app. Although this is standard research good practice, this was a sentiment voiced by many patients. The patient information sheet also included information that their feedback will be anonymous and only summary data will be reported back to practices. Advice was developed also for clinicians concerning when to ask about completion of the app within the consultation process. This was in response to patients' views that previous experience of data collection had focussed on the data collection disrupting the flow of the of the consultation process.

Future work could explore the experiences of patients who have used the data capture system and any challenges they faced with the operation of the system; this would allow further refinement of the system. It would be more significant to learn if users of the system felt this experience had an effect on their reflections about their symptoms and how they are managed. The notable gap in the literature is examining the effect of PROMs data collection and its impact on clinical care especially in patients with chronic conditions. Although the data capture system will identify patients' evaluations, it should be remembered that evaluations are patients' reactions to their experience of healthcare *e.g.* whether the process or outcome of their care was good or bad (Pascoe, 1983). Non-responders to such enquiries about preference and evaluations are more likely to be ill, less satisfied with the care provided, and less frequent users of healthcare than responders

although this is not always the case (Rubin, 1990). Future research will need to assess the level of non-completion of the app by patients to identify common features in an osteopathic setting. Any trends within non-participants will indicate the need to identify other means of gathering feedback from particular patient populations.

2.4.5 Conclusions

This qualitative study has highlighted a number of issues for careful consideration as part of the professional development for osteopaths as individuals, and osteopathy as a profession. The contribution of patients' views has been extremely valuable in the development of an electronic data capture system. First and foremost it provided an indication of whether patients would be willing to participate in practice-based data collection, and if an electronic system would be viable. Inevitably there were a range of views about this but few patients indicated inability or unwillingness to participate. The importance of collecting the "right type of information" was stressed, and asking patients their views on the most suitable PROMs to use was illustrative in terms of content and format. Further information was provided also concerning supplementary questions to be included to reflect issues that are important to them.

The views of clinicians (osteopaths, chiropractors, and physiotherapists) will be explored in the next chapter, and their contrast with the views of patients will be explored.

3

A qualitative study – investigating clinicians' views concerning the use of Patient Reported Outcome Measures (PROMs).

3.1 Introduction

Whenever a new initiative is planned in clinical practice it is wise to consult and engage those individuals who will be involved in such activities. Although the benefits of practice-based data collection, and the opportunities potentially available through the Any Qualified Provider (AQP) system have been well-documented, such changes could be uncomfortable for some clinicians particularly if they have not been part of their training or recent practice (Velicova et al, 2004; Hatfield *et al.,* 2007; Jette *et al.,* 2009; Swinkels *et al.,* 2011; Valier *et al.,* 2014). When the development of a Patient Reported Outcome Measure (PROM) data collection app was being planned, it was an ideal opportunity to try to engage osteopaths but also identify what would be feasible for them to use, what data would be useful to them and their practices, what support they felt they might need

with the process, and what fears or concerns, if any, they held. This qualitative work was extended to include members of the chiropractic and physiotherapy profession, and to learn from their experiences of using PROMs in paper-based formats or within electronic data capture systems.

The therapeutic relationship is based upon many factors, and some aspects of this relationship can be understood within the context of social exchange. Social exchange theory (SET) is among one of the most influential conceptual paradigms when understanding relationships in a range of different settings. It bridges a range of disciplines including anthropology, social psychology, and sociology. Even though different views of SET exist, theorists agree that it involves a series of, often interdependent, interactions that generate some form of obligation (Cropanzo and Mitchell, 2005). These interdependent transactions are regarded as having the potential to generate high quality relationships but this is contingent on the context of the transaction. Cropanzo and Mitchell state that "one of the basic tenets of SET is that relationships evolve over time into trusting, loyal, and mutual commitments" (Cropanzo and Mitchell, 2005). This occurs based on the observance of certain "rules of exchange" which have been described by Emerson as forming "a normative definition of the situation that forms among or is adopted by the participants in an exchange relation" (Emerson, 1976). In this manner the rules and norms of exchange can be regarded as guidelines informing the exchange process and can include the notion of reciprocity as one of the best known exchange rules. Within the realm of reciprocity, three different types have been distinguished including:

- Reciprocity as a moral norm ~ underpinning this type of reciprocity is the belief that this is based on a recognised standard of behaviour among individuals, and they should behave accordingly as a form of cultural mandate (Mauss, 1967; Wang *et al.,* 2003);
- Reciprocity as a transactional pattern of interdependent exchanges ~ outcomes are based on a combination of the efforts of both parties involved in the exchange (Molm, 1994; Molm, 2003);
- Reciprocity as a folk belief ~ including the cultural belief that people get what they deserve (Malinowski, 1932; Gouldner, 1960).

Notwithstanding the different types of reciprocity which have been distinguished, the opportunity to negotiate rules of exchange also exists to obtain the most equitable solution for both parties (Cook and Emerson, 1978). This can involve a more detailed level of negotiation and understanding, and are often present in economic as well as social exchanges. The difference between reciprocity and negotiation has been described extensively in the literature. Molm identified that "generally, reciprocity produces better work relationships than negotiations and allow for individuals to be more trusting of, and committed to, one another" (Molm, 2000), while negotiation is regarded as "inciting more unhelpful power use and inequality" (Molm, 1997).

The concepts of exchange of goods and services *e.g*. treatment for a fee is one of the foundations of healthcare in private practice settings. The delivery of care is based upon standards pertinent to that particular professional group, and based within common law (The Osteopaths' Act, 1993; GOsC, 2012). Within the therapeutic setting, there are also a considerable number of other factors that are relevant to a successful clinical encounter including the exchange of information, technical aspects of care, and service delivery (Kravitz *et al.,* 1997; Myers *et al.,* 2008; Georgy *et al.,* 2013; Leach *et al.,* 2013). The role of trust in the therapeutic encounter is based on the meeting of such expectations, and such trust can be underpinned by the existence of a parentalistic approach, or greater levels of patient engagement in clinical decision-making (Coulter, 1999; Hudon *et al.,* 2015; Fu *et al.,* 2016; Taylor *et al.,* 2016). These aspects of the therapeutic encounter are some of the many issues explored within this qualitative study. These are explained more fully in section 3.1.1.

3.1.1 Aims of the study

The aims of this study were:

- To identify osteopaths', chiropractors', and physiotherapists' views on the principle of practice-based data collection;
- To identify perceived practical issues including barriers and concerns;
- To identify information regarded as useful to collect during the process;

- To identify the information clinicians believe would be useful to be reported back to them, and in what format;
- To identify what support would be needed for practice-based data collection;
- To identify lessons learned from other clinicians who have been engaged with this process in other settings.

3.2 Methods

I used a qualitative phenomenological approach to explore clinicians' views about PROMs. A larger number of participants are involved than with the qualitative study involving patient participants, so a quantitative approach, *e.g.* a survey, could have been undertaken, but this would have potentially limited the depth of information obtained from the clinicians involved. Once again, guidelines by Tong *et al.,* 2007 have been used to guide the reporting of this study.

To meet the aims of this part of the study, a series of focus groups and individual interviews were undertaken. This process was undertaken with osteopaths, and then widened to include other clinicians treating patients with manual therapies (*i.e.* chiropractors and physiotherapists). To this end, contact was made with the Royal College of Chiropractors, and different groups and associations within the physiotherapy profession. Different aspects of the delivery of the study are described in the relevant section.

In contrast to Chapter 2, a combination of interviews and focus groups were used. The term focus group originated in market research but is used increasingly in social research settings (Merton *et al.,* 1956; Fontana and Frey, 2000). It is described also as a group interview or group discussion and this conveys some sense of what the process involves. The focus group, is essentially a group engaged in a discussion concerning a particular topic of interest; from this process data are generated (Ritchie and Lewis, 2012). The focus groups were used predominantly with osteopaths involved with the exception of 2 osteopaths who were unable to attend at specific times. Individual telephone interviews were used with chiropractic and physiotherapy participants for practical reasons as they were in

geographically diverse settings. Each approach has its own merits and disadvantages. In focus groups a type of "group bias" can be introduced when stronger personalities attempt to dominate the group, and influence other participants, especially if the group is very homogeneous (Wilkinson, 1998; McGee, 1999; Hughes and DuMont, 2002). Individual interviews can be arranged at more diverse times to accommodate work, domestic, and social needs but using telephone interviews can reduce the ability to capture subtle nuances that may accompany an individual's contribution (Carr and Worth, 2001; Shuy, 2003; Hiller and DiLuzio, 2004). Although comparisons between the two approaches are very limited, there is a more generalised view that both approaches have their merits, and the contributed data are still of high quality (Tausig and Freeman, 1988; Sturges and Hanrahan, 2004).

3.2.1 Ethics and research governance

This qualitative study was closely aligned to the study involving patient participants. It was clear that this was a research study as opposed to a different methodological approach. In common with the patients' study, ethical issues were considered including the rights and wellbeing of all of the participants, and the study was conducted according to Good Clinical Practice (GCP) Research Governance Framework (HRA, 2005).

The individual interview approach aimed to explore clinicians' views in great depth concerning the use and perhaps experience of using PROMs in clinical practice. There were no substantial changes to the questions used in the topic guide for patients, but the questions were reviewed again to ensure there was no potential to cause upset or distress to any clinicians. Modifications were made to the participant information sheets (PIS) to make it clear whether the qualitative study would involve an interview or focus group for each of the professional groups (Appendix 3.1). The PIS made it clear that all of the qualitative content would be anonymised, quotes would not be attributed to anyone in an identifiable manner, how the data would be used, and that participants could withdraw from the process at any time if they wished. This was to reassure all of the professionals that they could be as forthright as they wished in their views about using PROMs in a clinical

setting, and attributed data would not be shared. Participants were reassured that quotations would be coded during data analysis, and would be known to the research team only. Additionally, it was stated that although all audio-recordings would be transcribed verbatim, the transcribed data would be stored securely for a period of 20 years accessible only by the research team during that period.

A consent form was included with the PIS (Appendix 3.2), and this was returned prior to participation in the focus group or interview to ensure the participants' willingness to participate, and ensure that no questions were outstanding. Ethics approval for the osteopaths' focus group and interview study was obtained from the research ethics committee at Queen Mary University of London (QMREC1207), amendments were sought and obtained from the same committee for the studies involving chiropractors and physiotherapists (QMREC1207 – amendments 01-08-2014). Clinicians working in private practice only were included in the study; the ethics permission did not extend to NHS clinicians while working on NHS sites. When undertaking the interviews or focus groups, a common topic guide was used for all professional groups: this is described in the next section.

3.2.2 The topic guide

I described in section 2.2.3 the development of the original topic guide. Minor amendments were made to the topic guide used for patients including the removal of the questions relating to the example PROMs given to patients for discussion. Additional questions were included relating to potential support for clinicians when using PROMs. Once again, the topic guide contained a series of themes and broad questions to explore during the focus groups and interview process to understand how clinicians felt about the concept of PROM data collection in private practice. The key areas for investigation were:

- To identify current awareness and experience of using PROMs;
- To understand how participants felt about data collection in practice and if they had any concerns;
- What barriers, if any, were perceived in using outcome measures as part of your day-to-day practice?

- What support, if any, were needed to participate in collecting data and using outcome measures?

The complete version of the questions for the interviews and focus groups are shown in Appendix 3.3.

3.2.3 Piloting

Since the topic guide had undergone minor amendments to that used in the patient interviews, a pilot telephone interview was undertaken with a colleague to ensure that all of the questions were unambiguous, and were relevant to a clinical setting. I asked for feedback concerning whether any expected questions were absent; nothing was identified so the topic guide did not undergo further amendment.

While one of my main aims in this PhD was to develop a data collection facility for osteopaths in private practice, a considerable amount can be learned from those who have used PROMs in a range of different settings, and using other data collection systems. Interviews and focus groups involved osteopaths, chiropractors, and physiotherapists, and the sampling and recruitment for each professional group will be described in turn in the next section.

3.2.4 Sampling and recruitment - osteopaths

The sample of clinicians in this part of the study was a convenience sample drawn from osteopaths registered with the General Osteopathic Council (GOsC) Register of Osteopaths. A convenience sample is a type of non-random sampling also described as haphazard sampling (Etikan *et al.,* 2016). Members of the target population meet desired criteria which may be based upon accessibility, geographic proximity or ease of travel, and availability at particular times of the day (Dörnyei, 2007; Etikan *et al.,* 2016).

A convenience sample was drawn from osteopaths attending research hub meetings run by the National Council for Osteopathic Research, Regional Osteopathic Society members, and colleagues in local areas where the focus groups

were taking place. If osteopaths were unable to attend the focus groups, individual telephone interviews were held.

3.2.5 Sampling and recruitment - chiropractors

Osteopaths share many common areas of practice with their colleagues in chiropractic and physiotherapy. It has been documented that chiropractors and physiotherapists have experience of using electronic data collection systems, and the opportunity to learn from their experiences was regarded as being helpful to the development of a similar system for osteopaths (Moore *et al.,* 2012; Field *et al.,* 2016). Some chiropractors have benefitted from using the Care Response system, and the aim was to interview chiropractors who had used this system, and others who had not.

Chiropractors were recruited through considerable assistance from the Royal College of Chiropractors (RCC). Personal contact was made with the President following an introduction through a chiropractic researcher with a shared interest in PROMs (Jonathan Field). A recruitment email was created and circulated to members of the RCC.

3.2.6 Sampling and recruitment - physiotherapists

In contrast, recruitment of physiotherapists was considerably more problematic. A range of different initiatives (Appendix 3.4) was undertaken to try and explore as many potential avenues as possible. Contact was made with osteopaths and chiropractors who took part in interviews who worked in multidisciplinary practices to see if they could increase recruitment through personal networks also.

3.2.7 Participant focus groups and interviews

A series of focus groups for osteopaths were organised at a convenient time for participants to attend. Participants were advised they would last from 60-90 minutes. Interviews were arranged for osteopaths unable to attend other meetings, chiropractors, and physiotherapists around the country: these were held prior to clinic starting, at a convenient time during the working day, or at lunchtime depending on the preference of the interviewee. Participants in the telephone

interviews were advised these would last between 30-40 minutes. All focus groups and interviews were audio-recorded with prior knowledge and agreement of participants. The focus groups and interviews were designed to gain a wide geographic spread, and include clinicians with diverse years of experience, clinic settings, age, ethnicity, and sex. All participants gave signed informed consent prior to the interviews and focus groups taking place.

3.2.8 Analysis

All of the telephone interviews and focus groups were transcribed verbatim by me. Additional field notes were taken to record any significant events or comments associated with the focus groups and interviews. All transcripts were anonymised, and page and line numbers were inserted to facilitate ease of data handling, and make identification of references easier when undertaking analysis. A random selection of participants was asked to check their transcripts for accuracy to ensure that the content reflected their views in a process termed "member checking". This provided the opportunity also for participants to withdraw any statements with which they felt uncomfortable in hindsight (Guba, 1981).

Although time-consuming, this process allowed me to become immersed in the data, and allow thorough assessment of the content of the data after each focus group or interview. This gave me the opportunity to identify if any refinements were required in the Topic Guide in successive interviews or focus groups to phrase a question slightly differently to add clarity or to add some further dimensions to improve the depth of the response.

As described in Chapter 2, Pope *et al.,* 2001 indicated five stages in the analytical process when using the "Framework approach" (Ritchie and Lewis, 2012). This is summarised in Figure 3.1.

The Framework approach to analysis was employed again, and the initial phase of the analysis involved reading and re-reading the text of the transcripts and any field notes taken to becoming familiar with the content of each set of transcripts relating to the individual professional groups. After familiarisation, the transcripts were

assessed to see if any themes and topics began to emerge from the interviews. Emerging topics were then placed in categories, and then further placed under common themes. Field notes taken during the focus groups and interviews were examined also to explore if they added any additional context to the interview data.

Familiarisation

↓

Identifying a thematic framework

↓

Indexing

↓

Charting

↓

Interpretation

↓

Creation of matrices

Figure 3.1 A summary of the stages involved in the Framework Approach to qualitative analysis

3.2.9 Triangulation of data

After initial identification of themes emerged following immersion in the data, independent theme identification took place by DC and CF (DC* and CF†). Stage two of the data analysis involved discussion of those independent findings and the creation of an initial draft framework. This draft framework was then tested against a sample of transcripts to explore whether it was workable encompassing all of the themes identified or if further revision was required. Any disagreements could have been resolved by my second supervisor (RF‡), but this was not necessary. Once the final framework was agreed, quotes from the transcripts were organised and codes under appropriate themes and sub-themes.

DC* - Dr Dawn Carnes; CF† - Carol Fawkes; RF‡: Dr Robert Froud

Notes were kept on the decision-making process to ensure transparency of the analytical process, and observe the principles of good research practice. This process occurred for each individual professional group, and the frameworks based on analyses for each group are presented separately in section 3.3.

3.2.10 Organising and managing the data

The interview transcripts for each professional group were initially handled separately. They were allocated a reference code number to prevent either group or individual identification preserving the anonymity of all participants. Transcript pages were numbered, and the index process was undertaken by numbering each line of the transcripts making referencing easier.

Topics emerged from each group, and were progressively organised into themes in a systematic manner. References were then organised into a framework in Excel also using headings according to participant characteristics. This allowed data to be examined according to particular characteristics and explore any particular associations within the data.

3.3 Results

Three different professional groups participated in interviews and focus groups; their characteristics will be described in turn. The aim of recruitment was to include as diverse a sample of clinicians as possible in terms of geographic location, age, sex, experience, and clinical setting. The clinician participants were recruited from a wide geographical area. This is shown in Figure 3.2.

This section of the chapter will describe the characteristics of each of the participant groups, the frameworks which emerged from the focus groups and/or interviews, and the resulting themes, sub-themes, and models derived from those frameworks. Each aspect will be described in a separate sub-section.

3.3.1 Osteopathic population sample and characteristics

I undertook four focus groups and two individual interviews for osteopaths who were unable to join the focus groups between November, 2013 and June, 2014. In

total, 32 osteopaths were involved, and participants were based in regions around Exeter, Bristol, Haywards Heath, Leeds, Sheffield, and Newcastle-upon-Tyne.

Figure 3.2. Geographic spread of clinician participants in interviews and focus groups.

The focus groups were held either after clinic in the evenings (Bristol, Haywards Heath, and Leeds), Saturday morning (Exeter), or at suitable time between patients for the individual interviews (Sheffield and Newcastle-upon-Tyne).

The population sample of 32 osteopaths consisted of 13 men and 19 women; participants worked in a range of settings including as principals in single-handed practices, principals in group practices, assistants in group practices, in osteopathic educational institutions, part time in NHS settings, and as students in osteopathic education. The mean time since graduation was 14.2 years, and this was reflected in the age distribution of the participants with 21-30 (n =6), 31-40 (n=5), 41-50 (n=7), 51-60 (n=9), and 61-70 (n=5). Participants were predominantly white British (n=30) with two participants described as white "other". The full characteristics of all participants in the various regions involved are presented in Table 3.1.

3.3.2 Emergent themes and sub-themes - osteopaths

For each of the focus groups and interview participants, emergent themes and sub-themes emerged. These are described in Table 3.2. Data were organised into five separate themes describing different aspects of data collection, its potential impact on the consultation and practice visit, its impact on clinical management, bureaucratic considerations, and the knowledge and skills that might be needed to implement PROMs' use in practice.

Table 3.1 Full characteristics of osteopathic focus group and interview participants

Region	Age band					Sex		Ethnic group				Work								Years since graduation				
	21-30	31-40	41-50	51-60	61-70	M	F	White British	British Asian	White other	PSHP	PGP	AGP	PGP + NHS	AGP + NHS	PGP + Ed	Student	AGP + Ed	NP	0-10	11-20	21-30	31-40	41-50
Bristol (n=9)	2	2	3	1	1	4	5	9			1	3	4		1					4	1	3	1	
Leeds (n=6)	3		1	1	1	2	4	6				1				2	2	1		3		2	2	
Exeter (n=8)			2	4	2	6	2	7		1		6		2							1	6	1	
Haywards Heath (n=7)	1	3	1	1	1	1	6	6		1	1	2	4							4	1	1	1	
North East (n=1)				1			1	1			1												1	
Yorkshire (n=1)				1			1	1			1												1	

PSHP: Principal in single-handed practice
PGP: Principal in group practice
AGP + NHS: Assistant in group practice + NHS practice
PGP + Ed: Principal in group practice + working in OEI
NP: non-practising

Table 3.2. Analytical framework from osteopathic focus groups and interviews

Theme	Sub theme
1. Patient issues/engagement	1.1 Confidentiality and use of data 1.2 Relevance and appropriateness to patients (including simplicity) 1.3 Patient burden including parentalism
2. Clinical engagement	2.1 Confidentiality 2.2 Security of data 2.3 Cost/ resources
3. Clinical practice	3.1 Implementation into practice 3.2 Independent evaluation 3.3 Therapeutic relationships
4. Knowledge and skills	4.1 Uncertainty/lack of confidence 4.2 Analyses of data 4.3 Training (including accessibility and convenience)
5. Purpose	5.1 Comparative data 5.2 Language consistency/comprehension 5.3 Reflection/discussion 5.4 Dissemination

The themes and sub-themes with example data are described in the next section.

3.3.3 Osteopaths' qualitative data

Theme 1. Patient issues/engagement

1.1 Confidentiality and use of data

Participants in the focus groups wanted assurances on behalf of their patients that data would be anonymised, and not made available to commercial or other groups *e.g.* the regulator. Wider sharing of NHS patient data was a news issue at the time so this item was of particular concern.

"I suppose I'd just want to know that all of this stuff is anonymous. I'm a bit of a luddite with governance but I'd want to reassure my patients". (RA, lines 109-111).

1.2 Relevance and appropriateness to patients (including simplicity)

Some osteopaths expressed the view that too many patients, especially older patients, would be largely incapable of using a technology-based system, and a paper alternative should be provided. Others felt that this would be less of an issue. A small number of osteopaths stated that none of the PROMs they had seen so far were capable of capturing the outcome of an osteopathic consultation.

"The disadvantage for me is how you transfer something that is used for pain to reflect all of the types of symptoms that patients come into practice with". EE, lines 8-9.

"In the NHS this is what they're trying to do and I mean patients of all ages are used to filling in the touch screens when they go into the practice to say that they are there so there's some level of skill at all ages just as long as it isn't too complicated to negotiate". EB, lines 281-284.

"We have a lot of older patients and they would be nervous … they would be very afraid of breaking the computer". HB, lines 168-9.

"But what would be better if we could just find an outcome measure that would measure all of those things and would be holistic enough to be relevant to osteopathy. It is important that any measure we use can reflect what we do as osteopaths and osteopathic thinking which is more holistic than other professions". LC, lines 154-157.

1.3 Patient burden including paternalism/maternalism

Concern was expressed by some osteopaths that patients were attending appointments within an already busy and demanding life and this request would be regarded as an added burden.

"It can be a challenge to get the patients to fill these in especially if they come in and they are very acute and they just want to get some treatment" EB, lines 17-18.

"Most patients really just want you to get on and treat them and they don't want to waste their half hour". HD, lines 102-103.

Theme 2. Clinical engagement

2.1 Confidentiality

Some osteopaths felt that the PROM data should be linked to some form of identifier that would allow their data to be tracked. This would allow the osteopath to use the PROM as a clinical monitoring tool.

"I'd want to be able to follow patients up. Presumably this would be possible. We wouldn't be looking at this as a one-off because that wouldn't be much good". HC, lines 84-86.

2.2 Security of data

Information about the storage of data, and the potential for hacking was discussed. Osteopaths felt patients would need clear statements about the security of their data.

"There's so much stuff around hacking, and the CareData stuff. I need to be able to tell my patients where the data is going and who will see it" BC, line 44-46.

2.3 Cost/ resources

The resource of most concern was practice time. This included that of the clinician having to discuss the PROM with the patient, and also the reception staff who would need to help patients with the PROMs completion, and dealing with questions. Questions were also asked about the cost to the practice of using the PROM on a regular basis.

"But there are problems with costs and some companies having a monopoly in terms of cost. For some practitioners they would see buying the app as a barrier and they may try and put patients off using it". EE, lines 359-361.

"I would be happy to pay for this to be done for me … something like £50 per year say which is pretty cheap to have his sort of data returned to you without the hassle of gathering it and analysing it yourself". BD, lines 205-7.

Theme 3. Clinical practice

3.1 Implementation into practice

Views were expressed that implementation should be across the profession. Some osteopaths felt they could potentially lose patients if they were asking them to complete questionnaires, and other practices were not. Others expressed a polarised view that it would be a disadvantage not to engage in the process. To others it was just another demand on their time.

"Quantifying what we do using good outcome measures is helpful because we mustn't forget that most osteopaths are in commercial competition with other manual therapy professionals" EA, lines 92-94.

"Bureaucracy… that's all it is bureaucracy … It drives me mad…having to give out bits of paper. I don't want patients turning up early so they're hanging around the practice and have to fill these things in, and then have to give them back to me. I don't really want to deal with them". LA, lines 72-75.

3.2 Independent evaluation

The fact that practice data could be evaluated independently was seen as a significant benefit from using PROMs. Other osteopaths made the point that requesting feedback is something very familiar today and it might appear unusual if osteopaths did not collect this type of data.

"I have found it quite helpful in the past mainly to be able to talk to patients and be able to give numbers to patients who have responded well and say about how much something has increased in a way that they understand" EE, lines 43-45.

3.3 Therapeutic relationships

Discomfort was noted that some patients who had been attending for a considerable time, and with whom some osteopaths had built strong therapeutic relations might feel uncertain about completing the PROM. This was largely a reflection on a PROM's ability to capture change in a patient with long term symptoms. Others noted that some patients might regard this as a natural progression in their care, as it is common practice in NHS settings.

"But I think patient reported outcomes aren't something that we should fear as we do pride ourselves on being patient-focussed" EC, lines 84-85.

"Yes… well you can get patients who simply don't like you... you can get others who will either be really satisfied or really unsatisfied (sic) and how will their data ... if the people in the middle don't participate but only the other groups do … I don't know…" HC, lines 239-242.

" I have had one patient whinge at me during a consultation to the effect that I came here to get my back fixed 'not to fill in bloody forms'". BB, lines 283-4.

Theme 4. Knowledge and skills

4.1 Uncertainty/lack of confidence

Some osteopaths stated that they had used PROMs before and had collected a considerable amount of information but been unsure how to analyse, interpret, and apply it. They felt there was nowhere to go to find out this type of information for a novice user.

"I'd be happy if I could have a computer programme and then I could have a go". HG, lines 344-5.

4.2 Analysis of data

Other osteopaths expressed the view that while they did not mind collecting data, they did not want to have the role of analysis, and would expect that to be done by someone else.

"I would be happy to let someone else do it. I really don't like statistics and numbers". HE, lines 340-341.

"I would prefer it just to be done by someone else and then tell me what they've found. At least I know it would be done right, and it's a lot of work". HC, lines 359-360.

4.3 Training (including accessibility and convenience)

When the issue of training was raised, osteopaths noted that an online resource would be useful. This would be preferable to having to make time to attend a course with travel and other costs this might entail *e.g.* missing time from practice.

"Some basic training because I've looked at them but I don't feel I'm quite sure about what I'm doing and whether I'm using them right". BE, lines 96-97.

Theme 5. Purpose

5.1 Comparative data

The view was strongly expressed that osteopaths did not want their individual outcome data to be shared in the form of a league table. Aggregate data across the profession was regarded as useful, and would act as a benchmark against which individual practices or clinicians could compare their own practice at a private level.

" I think there are potential fears. I don't personally have them but I can see that there would be fears especially if every practice has to collect this data then this would be something that would be very unsettling if it meant comparing one practice with another, and then if that has to be the way that you have to release your data like

a…hospital. Personally I think it would be good to have access to a report of your own data and be able to compare yourself to a national trend, but I don't think it would be good to be able to look at another practice's summary data". EG, lines 429-435.

5.2 Language consistency/comprehension

Osteopaths practising in inner cities expressed concern about the availability of PROMs in languages to suit all ethnic groups. Others felt that the language in some questionnaires they'd experienced was somewhat vague.

"It's the language we use... It predisposes to an episodic view of neck and back pain but it doesn't capture those people who do have a meandering course of symptoms". BB, lines 389-391.

"Where we work in our inner city practice, there are people like asylum seekers and other migrant populations, they would struggle with some of this stuff frankly". BC, lines 395-6.

5.3 Reflection/discussion

The value of comparative data was clear, but one concerned view was expressed that osteopaths might discover their practice was not as good as they had thought. Some osteopaths expressed the view that learning about negative feedback would be helpful.

"*But I think that is the reality of practice life that you will always get a small percentage that aren't happy and in many ways it's much better to know even though it might make uncomfortable reading"*. ED, lines 342-344.

"I think for us when I've used something like this is my practice looking at patient satisfaction it has been a way of reiterating how important patients' views are to us". EB, lines 533-534.

"But isn't it … well a problem that the people who are less likely to do this… less receptive are the ones who perhaps have more need to look at... at how they're doing". HB, lines 454-6.

"I'm going to say it…I might find out I'm no good… and I know I should want to know but I don't". BG, lines 161-2.

5.4 Dissemination of findings

Osteopaths wanted to know also how the data would be fed back to them, and gave indications of what would be the most useful as a resource for the practice to use. Osteopaths were also asked about the type of information they thought patients might want to know about their practices.

"Well you know I'm not sure… er I know this is supposed to be good practice but in my experience most patients don't really care". EC, lines 354-355.

3.3.4 Evolved models and their meaning – osteopaths

An extensive range of views on the topic of PROMs and their value to practice were supplied by the osteopaths participating in the focus groups and individual interviews. The values and drawbacks of using PROMs, the fact that their use might meet patients' expectations as patient feedback was required in more and more situations both clinical and non-clinical. There was a view that patients might be less interested in summary findings, but this might provide considerable insight for clinicians, albeit that some of it could be uncomfortable. For the osteopaths who could see the value of using PROMs, they were clear that the bureaucratic burden should not be too great, and they would not want to have the responsibility for collecting and analysing data. These views have been summarised in a model (Figure 3.3).

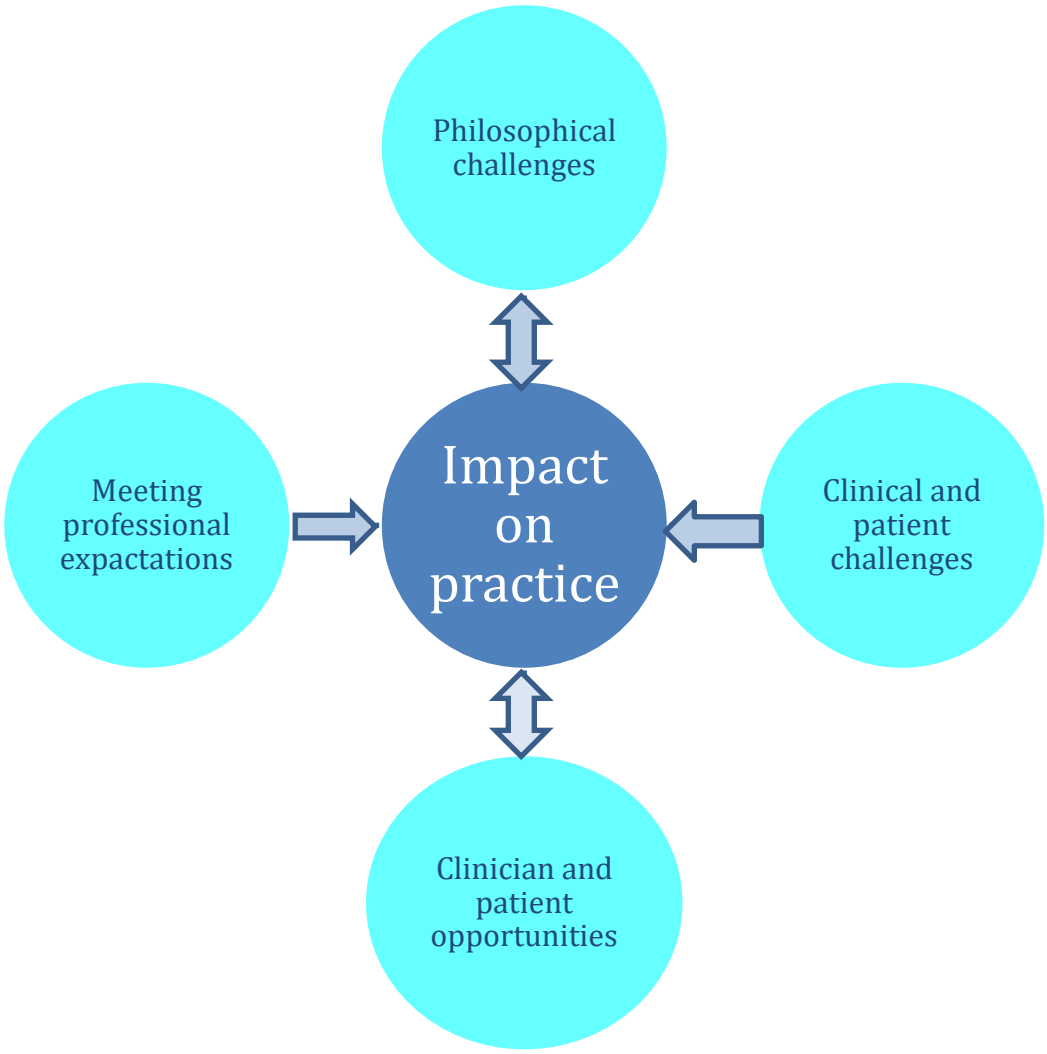


Figure 3.3. The potential impact of using PROMs in osteopathic practice.

3.3.5 Chiropractic population sample and characteristics

A total of twelve chiropractors responded to my recruitment email, and ten were ultimately recruited. Individual telephone interviews were arranged between September and December, 2014. The interview sample consisted of four men and six women; participants were working in a range of settings including as principals in group practices with other chiropractors only, clinicians from other disciplines, assistants in group practices, in chiropractic educational settings, one was on maternity leave, and one was no longer practising. The mean time since graduation was 16.2 years, and this was reflected in the age distribution of the participants with 21-30 (n=2), 31-40 (n=4), 41-50 (n=3), and 51-60 (n=1). Participants were

predominantly white British (n=10). The full characteristics of all chiropractic participants in the various regions involved are presented in Table 3.3.

3.3.6 Emergent themes and sub-themes - chiropractors

Emergent themes and sub-themes were identified from each of the individual chiropractic interviews. These are described in Table 3.4. Data were organised into five separate themes describing different aspects of the data collection process, its potential impact on patient care, its impact on clinical and practice management, and the potential for personal development that PROMs' use in practice could offer. The themes and sub-themes with example data are described in the next section.

Table 3.3 Full characteristics of chiropractic interview participants

Region	Age band					Sex		Ethnic group			Work								Years since graduation				
	21-30	31-40	41-50	51-60	61-70	M	F	White British	British Asian	White other	PSHP	PGP	AGP	AGP + NHS	PGP + Ed	Student	GP + Ed	NP	0-10	11-20	21-30	31-40	41-50
South West (N=2)	1	1				1	1	2			1							1					
London and South East (N=3)	1	1		1		2	1	3			2							1	1	1		1	
Hampshire (N=4)		2	2	1		1	3	4			3				1								
Yorkshire (N=1)				1			1	1			1										1		

PSHP: Principal in single-handed practice
PGP: Principal in group practice
AGP + NHS: Assistant in group practice + NHS practice
PGP + Ed: Principal in group practice + working in OEI
NP: non-practising

Table 3.4 Analytical framework from chiropractic interviews

Theme	Sub-themes
1. External engagement	1.1 Business development 1.2 Communication with other healthcare professionals 1.3 Philosophical and personal challenges 1.4 Using electronic resources 1.5 Motivation
2. Clinical competence in patient care	2.1 Demonstrating progress 2.2 Wider clinical issues 2.3 Patient follow up
3. Practice management	3.1 Use of practice resources 3.2 External support 3.3 Value of PROM data 3.4 Management appraisal 3.5 Communication about research
4. Patient feedback	4.1 Identifying practical issues 4.2 Identifying clinical issues 4.3 Supporting research
5. Personal development	5.1 Individual benefit and reflection 5.2 Professionalism 5.3 Marketing

3.3.7 Chiropractors' qualitative data

Theme 1. External engagement

1.1 Business development

Some of the participants mentioned the potential opportunities available from introduction of the Any Qualified Provider (AQP) system introduced after political changes in 2012. They noted that PROMs had a role to play in collecting data to facilitate the acquisition of contracts within this service.

"It was in our minds that we wanted a better feedback system and this just sort of turned up. I'm really pleased it came along." C3, lines 104-105.

"One thing that is in the back of my mind is GP commissioning and the opportunity to get chiropractic as part of NHS care. I think the PROMs system is very good to be able to show that we are monitoring patients properly. I think it will be a strong addition." C3, lines 108-111.

1.2 Communication with other healthcare professionals

The uniformity of PROM data was regarded as valuable when communicating with other clinicians. It was regarded as a form of common language, and the fact that the data came from the patient strengthened its appeal.

"We write to GPs but we are better at it since we've had the PROMs system. We like the fact that it uses very straightforward graphs and we've been using that to include in a GP letter when we have got a set of results and we can show them the patient's improvement." C3, lines 55-59.

1.3 Philosophical and personal challenges

Issues arose in some interviews concerning the lack of engagement of some clinicians, and the underpinning beliefs that motivated such issues. In other cases a lack of congruity about clinicians' place in the wider healthcare arena were noted.

"There is a section of the profession that is quite resistant to engaging with anything outside the profession. They don't see it as relevant. They are a very small section but they have no desire to work within the NHS, and they don't see talking to commissioners or other groups as part of their professional role at all. Actually they quite like working in isolation and just focussing on a few patients. I don't have a great deal of sympathy with that point of view." C2, lines 206-213.

"I think there are a number of reasons why people don't use them, I think that often they don't think they're relevant or important depending on what paradigm or philosophy they believe in, I've seen this in chiropractic and osteopathy where sole practitioners work within some rather strange paradigms, and where they feel a PROM wouldn't have any value to them." C7, lines 175-179.

1.4 Using electronic resources

Many of the clinicians interviewed had electronic systems working in their practices. A small amount offered paper questionnaires also, and recognised some of the limitations of using electronic resources for different patients.

"There are some patients who are not very good with their email, so we bought a tablet for the practice for patients who had email but didn't actually access it very well at home. But we find that using the tablet in the clinic is pretty onerous with patients who, for example, can't see very well so you've actually got to go through it all and explain it all to them. So buying a tablet has been a bit of a waste of time." C3, lines 155-160.

"When patients decline to use the electronic questionnaire, I think a lot of it is about how computer savvy they are… I think that's got a lot to do with it. I think the other thing, and I don't know about you, but I'm always very wary if someone wants me to sign up for something because the next thing is you're going to be bombarded with all sorts of emails and marketing things and as much as we don't do that I do understand people's reticence." C6, lines 124-128.

Theme 2. Clinical competence in patient care

2.1 Demonstrating progress

Clinicians noted that the PROMs were a useful adjunct to their day-to-day practice. On occasion issues had been raised when completing the PROM which had not naturally arisen within the consultation.

"As a clinician I want to know how my patient is doing not just how I think they are doing, but also think they help by showing respect for the patient's perspective and that you value that, and that is really helpful in that aspect of care." C2, lines 3-6.

"It has been useful and … also I would say… well it seems to have helped them [patients] and us with some issues that I might not have brought up especially when it concerns some of the psychosocial issues and activities of daily living. I think

sometimes we have a tendency to focus just on the pain, and if the pain is going in the right direction we must be doing the right thing… but actually it's important to cover those things about the activities of daily living and how patients actually feel about some things other than pain." C2, lines 102-106.

2.2 Wider clinical issues

Clinicians shared their experience of using a PROM data collection system noting that patients had subsequently shared information about other issues of importance to them which may have been directly or indirectly associated with their presenting symptoms.

"There may be something about which they're quite frightened and they would have difficulties voicing them face-to-face. And it may be that we see this and I think this is something we need to mention to the GP." C1 lines 33-37.

2.3 Patient follow-up

Some clinicians noted that they had changed their practice in following up patients after treatment after some had recorded worsening of symptoms after treatment. This extension of care was reported as being beneficial to patients for reassurance, and the opportunity to offer focussed advice.

"Other times they will have put on the form that their pain in their leg say is that bit worse after treatment, and perhaps their symptoms have deteriorated but they won't actually pick up the phone to come in so we have to say to them you know if you have pain and you are worried, or you feel things are getting worse then just give us a call and we can talk to you about it over the phone." C7,lines 50-54.

"I think it's good for patients to be able to have their say and especially when they don't have me sitting in front of them." C3, lines 3-4.

Theme 3. Practice management

3.1 Use of practice resources

While many clinicians were positive about using PROMs in clinical practice, they stated that key considerations for them had been the impact on time for them as clinicians and for their administrative staff, and the smooth running of the consultation process. Although some recognised initial reluctance in their staff, with a combination of gentle and not-so-gentle cajoling, PROMs had been introduced and been beneficial for the running of the practice and patient care in the longer term.

"I'd really tried to er look at er er different methods of such like [PROM data collection] before but it was just that this was one that looked after itself, ran itself, managed the data, erm and that was what did it because it was so easy, and there is a time factor in a busy, busy clinic like we are, and we wanted something that just didn't interfere." C1, lines 8-12.

"The biggest concern is the added load on the front desk staff and for clinicians… you know it's another thing to think about or do when you already do a lot of form filling and record-keeping and this has added to that. We keep trying to cut some of that and streamline things, but to get data you need to collect information and get something that is comprehensive." C2, lines 130-135.

3.2 External support

Some clinicians stated that they had been apprehensive about using an electronic system, and recognised that they may not been using their current system to its full potential. The availability of ongoing support was an important aspect to its introduction and success.

"It's the technology side of it that we will need help [with]. And I know when we need help we will get it as someone from the College of Chiropractors will come along and help us as they did when we were setting it up." C3, lines 147-151.

3.3 Value of PROM data

While some clinicians felt that being able to view whether you were performing to a standard comparable to your peers was important, they also felt that having a body of data about the chiropractic profession *en masse* was helpful in demonstrating what the profession could offer in terms of patient care.

"The main thing is that we can use this as a benchmark against national standards. We also like to think that the Royal College of Chiropractors can use the data with the things that they do, and show the things we are doing as part of the clinical management of patients." C1, lines 189-193.

However, there were some slight cautions to the use of PROMs. For some clinicians their inability to capture the range of symptoms commonly reported by many patients experiencing musculoskeletal symptoms was problematic. Many PROMs are disease specific or very generic, and their perceived failure to encompass many common symptoms while focussing on single symptoms was a cause for slight antipathy.

"The only other thing I would stress is that so many outcome measures focus on pain rather than more common symptoms like stiffness and aching, and they do seem to be based within diagnostic areas which I appreciate they have to be but that can be difficult in patients with multifocal symptoms." C2, lines 237-241.

3.4 Management appraisal

The value of PROM data extended beyond solely clinical performance for some chiropractors. One participant noted that PROMs had been useful as part of a management appraisal identifying areas of strong practice and results, while indicating others where some targeted Continuous Professional Development (CPD) activities had enhanced patient management. Using other measures of outcome, *e.g.* satisfaction and experience as part of a data capture system had been useful also when reflecting on individual performance as well as for third party appraisal.

"I think it does for say if I had an appraisal and if the chiropractors in the practice wanted to look at how my role as a chiropractor was going in their business I think it is invaluable for them to show that the people in their business are receiving the type of service they would like to receive but also it's a way of being self-reflective". C4, lines 231-235.

3.5 Communication about research

When asked what information they thought patients might want, clinicians were less certain this was needed. The value of day-to-day data collection was noted based upon a perception that this type of evidence was expected by some patients.

"I don't think it's very high on people's agenda [feedback about data]. They just want to get better. They don't really care what is happening to the data and how other patients are doing just as long as they get better themselves". C3, lines 141-143.

"I think there's a shift in …I think a side shift in GPs that PROMs are a force for good rather than traditional trials and that practice-based research is getting more recognition and more notice". C5, lines 86-89.

Theme 4. Patient feedback

Participants noted that giving patients the option to provide feedback in a manner remote to the practice had been helpful. This had allowed issues in clinical and practice management to be addressed.

4.1 Identifying practical issues

The issue of what happens to patients who no longer continue treatment was raised during the interviews. Having an electronic system to capture data from this patient group was regarded as helpful.

"There have been one or two patients who have shown a negative point of view who have perhaps decided not to continue with …er treatment, and sometimes they feel comfortable expressing it on paper using the Care Response system rather than saying it to me or perhaps the receptionists, or perhaps having to talk to someone on the

phone. And that has been quite helpful because sometimes it has genuinely been something wrong that we weren't aware of". C1, lines 69-75.

4.2 Identifying clinical issues

Being able to document change and compare change across patients was helpful for some clinicians. The value of examining outcome in a more systematic manner rather than relying on perception of change was instructive.

"I have been really surprised how it's shown patterns of improvement, and it just helps sharpen up what you do, and has really been incredibly useful". C1, lines 16-18.

4.3 Supporting research

While some participants felt that being able to contribute to research through using PROMs was important to them, they emphasised also the importance of fully informing the patient about the function of the research and its intended impact on practice. For other contributors using a tool that had a professional and clinical resonance helped to underpin their commitment to the data collection process.

"My preference from the start was for Bournemouth [questionnaire] as it had been designed by a chiropractor for our type of clinic". C1, lines 140-141.

"I didn't say it was a necessary thing but I always used to mention it on the first visit that it was something we were keen to do in the practice to make sure that we were evidence-based and best practice and that sort of thing, and I think if someone knows why they are filling in a form it's much better than if you're just shoved a form". C4, lines 67-71.

"So my concern is that people are not using them and that may be a reflection on the fact that I place too much importance on them and I've not looked at the potential limitations and down-sides, but I think that yes my main issue is that we're not using them that we're not willing to gather information from third parties enough". C8, lines 111-115.

5. Personal development

Although the role of PROMs and other data collection has been discussed for its value in research, patient management, and clinical management, the relevance for the individual was discussed by participants.

5.1 Individual benefit and reflection

The usefulness of comparing personal practice and performance was discussed by some contributors. While some participants had noted the use of PROMs data as part of their appraisal, for others it had value for their personal development as a clinician.

"It has made us much more aware of how we are doing compared with others, and it has made us much more critical of our own performance. You tend to think that everybody you treat is getting better but that's not always the case, and in some cases when patients don't come back you can't just assume that it's because they got better. So it has improved our critical analysis and has made us a lot more realistic I think. The outcomes may not always be what you assume and expect". C2, lines 113-120.

"One of the key things obviously in the relationship is communication and it's something I feel I am quite skilful so for me I had ...erm ... had expected it to be good but was really pleased to see that that was at a high level". C4, lines 79-82.

5.2 Professionalism

The fact that data are collected in so many different places now was mentioned by some participants. The significance of doing this within a healthcare setting was viewed as important, and in some views long overdue within chiropractic generally.

"I think it's vital for the profession. We have been very tardy in collecting data as a profession, and this slows us down in trying to make meaningful contact with policy makers who could make a difference for the profession… and for patients also. It is burdensome and a bit scary to collect it but you just have to get on with it". C2, lines 199-203.

"After all, it's not about how good you think you are but what the patients think about your treatment. That's the reality and that's what we like about the peer-review that sometimes you can get surprises in that you're not doing as well but if you don't face it you can't fix it". C1, lines 237-241.

5.3 Marketing

The interviews enquired about clinicians' perceptions of using PROMs data in marketing their practices. Although there has been a focus on patient and clinic management, the value of the data in marketing was less frequently discussed but nonetheless was regarded as a helpful discussion tool.

"If I am trying to engage a third party or anyone else who is thinking about paying for my services whether that's an insurer or the NHS, they are really interested in outcome measures. They're just part of the healthcare landscape now". C2, lines 7-10.

An extensive range of views on the topic of PROMs, and their value in practice have been contributed by chiropractic participants. These views have been summarised in the next section.

3.3.8 Evolved models and their meaning - chiropractors

It was clear from the interviews conducted with the chiropractic participants that the use of data capture systems and PROMs was more familiar to them than for the osteopathic participants. Using PROMs and collecting other data concerning experience and satisfaction helped to highlight for them whether patients were responding to treatment as anticipated. This provided the opportunity for patients to highlight other clinical issues that might be of concern that had not been raised within the consultation. The ability to be reflexive based on such feedback to either change a treatment plan or refer on to another healthcare professional promoted a sense of individual agency for patients and clinicians respectively. This also raised comments about the importance of patient feedback in dealing with practice management issues *e.g.* following up patients after treatment, and dealing with

patients' opinions about issues in the practice which were important for the practice management to learn about and thereby respond.

It was noted that receiving PROMs' feedback can be a slightly daunting proposition for some clinicians especially if the patients' responses were not as good as anticipated. While occasionally uncomfortable, it was recognised that this was an important learning opportunity to allow practise to be changed by reflection or additional training as appropriate. The importance of feeling part of a professional collective was also noted by some participants. While PROM data represented potential individual opportunities, pan-professional data was regarded as important to support discussion about the role chiropractic could offer within general healthcare delivery. These views have been summarised in a model (Figure 3.4).

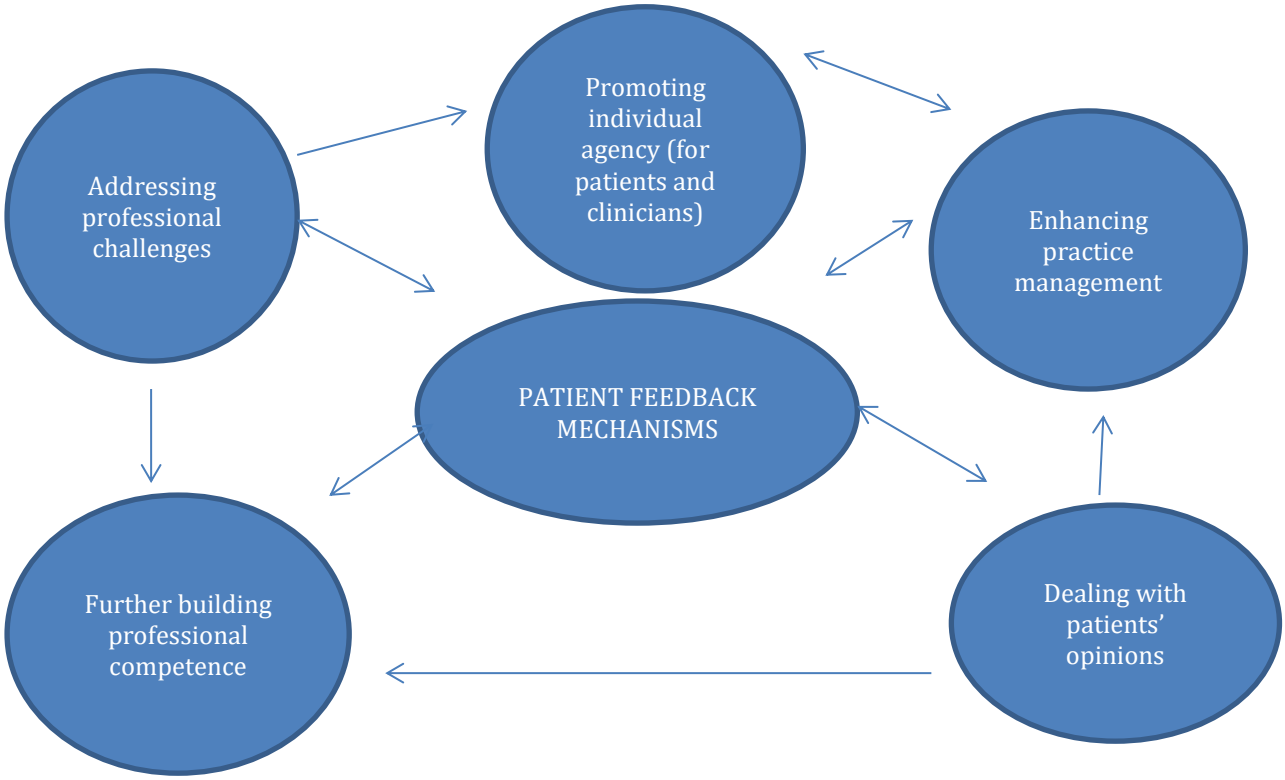


Figure 3.4 Chiropractic model developed from themes and sub-themes.

3.3.9 Physiotherapy population sample and characteristics

Recruitment of physiotherapists occurred through a range of different methods as described in section 3.2.6. The final level of recruitment was both disappointing

and frustrating. I undertook four individual interviews with physiotherapists (two men and two women). Participants were working in a range of settings including as principal in a group practice, part time in NHS settings and private settings, and part time as a student and in clinical practice. The mean time since graduation was 27 years, with the age distribution of the participants being 31-40 (n=1), 41-50 (n=1), and 51-60 (n=2). Participants were varied in their ethnicity including two participants who were white British, one white "other", and one self-described as British Asian. The full characteristics of all participants in the various regions involved are presented in Table 3.5.

Table 3.5 Full characteristics of physiotherapy interview participants

Region	Age band					Sex		Ethnic group			Work								Years since graduation				
	21-30	31-40	41-50	51-60	61-70	M	F	White British	British Asian	White other	PSHP	PGP	AGP	AGP + NHS	PGP + Ed	Student	AGP + Ed	NP	0-10	11-20	21-30	31-40	41-50
North West (n=2)		1	1			1	1	1	1			1		1							1	1	
Midlands(n=2)		1		1		1	1	1		1				1		1			1			1	

PSHP: Principal in single-handed practice

PGP: Principal in group practice

AGP + NHS: Assistant in group practice + NHS practice

PGP + Ed: Principal in group practice + working in OEI

NP: non-practising

3.3.10 Emergent themes and sub-themes – physiotherapists

Emergent themes and sub-themes were identified from each of the individual physiotherapy interviews. These are described in an analytical framework in Table 3.6. Data were organised into five separate themes describing the perceived burdens and benefits of data collection process, how data collection is regarded as a professional requirement for some, the issues associated with implementation, and the manner in which that implementation can affect patient engagement. Finally, factors surrounding the appropriateness of measurement were discussed.

Table 3.6 Analytical framework from physiotherapy interviews

Themes	Sub-themes
1. Burden and benefits	1a. Keeping up to date with new initiatives. 1b. Information and advice for business development and patients. 1c. Reflection on practice. 1d. Added bureaucracy for patients and clinicians.
2. Professional requirements	2a. Dealing with the demands of external agencies. 2b. Demonstrating effective practice. 2c. Demonstrating efficiency of practice. 2d. Challenges to existing skills.
3. Patient involvement	3a. Engaging patients in the treatment process. 3b. Identifying what is important to the patient experience. 3c. Evaluating the service delivered. 3d. Supporting conversations with patients about outcome.
4. Implementation	4a. Coping with different levels of literacy and language. 4b. Ensuring accessibility for everyone in the population. 4c. Ensuring data returns. 4d. Organisation and costs associated with PROMs. 4e. Standardisation of measurement instruments
5. Appropriateness of measurement	5a. Identifying instruments sensitive enough to measure change. 5b. Being aware of the necessary limitations of measurement. 5c. Being clear about the purpose of measurement.

3.3.11 – Physiotherapists' qualitative data

The themes and sub-themes are expanded with accompanying quotes in this section.

Theme 1. Burden and benefits

Physiotherapy participants were more familiar with mandatory use of PROMs in NHS settings, and in private practice at the request of insurers. They highlighted that were both burden and benefits to this.

1a. Keeping up to date with new initiatives

The importance of considering new innovations in practice was discussed by one participant. The change from paper-based to electronic practice management was challenging to implement.

"We moved from a paper-based system to a computerised practice-management system about 10 years ago. At the time there was some muttering [from some clinicians and staff] but they were dragged into the nineties and it has been effective". A, 198-201.

1b. Information and advice for business development and patients

The usefulness of PROMs in highlighting clinical issues that were important for patients was noted. This could allow early management of issues thereby preventing use of additional services delivering a cost saving.

"I suppose the main thing is to think about how much benefit an actual service adds to patients. I imagine that's what you're thinking about but it's also a case of showing how much is added and how little other services have to be accessed as a… a result, or what other added value is part of the service". B, 192-201.

1c. Reflection on practice

Feedback about care delivery and outcome was viewed constructively producing an impetus for action.

"The sort of positive comments from clinicians is that they have regarded it as an opportunity for self-reflection and they have found that if feedback has come back which wasn't quite as great as they were expecting then they knew they needed to do something about it". D, 191-195.

1d. Added bureaucracy for patients and clinicians

Some completion of paperwork is necessary in various practice settings, and PROM completion was regarded as a potentially unwelcome addition to existing requirements.

"Initially there was some reluctance because there is a lot of paperwork to get through within a short space of time with patients". D47-49.

Theme 2. Professional requirements

In some clinical settings the completion of PROMs was mandatory for clinicians, and this was discussed in the context of various care delivery settings.

2a. Dealing with the demands of external agencies

The demands for completion of PROMs came not just from management tiers within a clinical setting, but from external regulatory organisations.

"There is a lot of important work going on at the moment, and safety and governance and all of the things that the CQC come and look at, and there's just so much to do. We can use patient-focussed measures so that we can show we are being patient-focussed". C, 150-153.

2b. Demonstrating effective practice

Clinicians in both private practice and from experience in NHS settings felt that completion of PROMs to demonstrate effective practice was an accepted part of publicly-funded healthcare and dealing with external funding agencies.

"It's the case already in private practice that we're dealing with insurance companies and they're all pushing down the same pathways on effective practice and proving that your practice is effective you know". A, 45-49.

2c. Demonstrating efficiency of practice

Economic pressures in publicly- and privately-funded healthcare underpinned the idea that care delivery should be as efficient as possible.

"I think it's a huge opportunity, auditing your practice for research, standardising things, selling your practice to commissioners that sort of thing, to be in that kind of position overall I think it's about improvement or hopefully it is". C208-211.

2d. Challenges to existing skills

Clinicians who had been in practice for an extended period of time recognised that skills must be kept up-to-date; this included technical skills and other aspects of patient management *e.g.* PROMs use.

"When you are in private practice you can get a little bit isolated and stuck in a system which only you use and you can use it quite happily ad get stuck in a rut but then something comes along like PROMs and it can be a bit unnerving". A, 24-27.

Theme 3. Patient involvement

The manner in which PROMs' use could have a beneficial effect on patient care was discussed.

3a Engaging patients in the treatment process

Although it has been suggested by participants that PROMs can benefit clinicians; their use in getting patients more closely involved with the therapeutic process was also highlighted.

"It should present opportunities to reflect, and hopefully improve I would have thought yes and engage the patients as well". C202-205.

3b. Identifying what is important to the patient experience

The opportunity to learn from patient feedback, and being responsive to things that were highlighted by them was noted as an important part of using measures of outcome *e.g.* experience.

"We use a questionnaire that has about ten questions on it [experience] ranging from how did you find the experience, booking an appointment, what did you think about the physio that treated you, what about the information provided that sort of thing and that's our kind of basic data collection; and we just do a sample of patients every month. And then I'll sit down and read through them and if there are common themes that keep coming through then I can address them". A, 171-178.

3c. Evaluating the service delivered

The differences between the technical aspects of care, and service delivery were noted by some participants.

"Obviously it's important for the physiotherapists because they need to know about what is making a difference from the perspective of delivering the service and delivering the different interventions. But maybe the difference at the service level is most important rather than looking at specific treatments". B, 28-32.

3d. Supporting conversations with patients about outcome

The opportunity to monitor response to treatment, and use this as an objective rather than anecdotal measure was highlighted. This could offer reassurance to some patients who had forgotten their health and/or disability status at the onset of treatment.

"I think they think of them from the clinicians' point of view that they can have that conversation with the patient if the patient doesn't feel that they are improving hugely then they can look at the patient's specific functional scale score when they first started and shown them the latest one and say well when you first started you were this and now you are this so you are indicating a change. You know so it does help to reinforce that so it can be useful for some patients basically". D 165-172.

Theme 4. Implementation

4a. Coping with different levels of literacy and language

The need to make PROMs accessible to all patients especially in areas with diverse ethnic populations was cited.

"We have a lot of people who may speak English and get by with that but can't read it, so we have a very diverse...very diverse population, and I think I read that the average reading age is not what you would think it would be in an adult population as well , erm so I think there are practicalities there mainly focussing on the language but maybe also on other issues". C111-117.

4b. Ensuring accessibility for everyone in the population

Ensuring access for IT-literate, and non-IT literate individuals, or those without IT access was discussed by some participants.

"I suppose language-wise that would still be a problem completing it online or on paper. As far as technology goes, I don't know what...whether people have access to the internet but I assume not everybody does, erm and the same with a smartphone I'm not sure everybody does, so in terms of using the internet not everybody does, and it could exclude some of the elderly and maybe we would have to offer a choice of completion methods, and that sounds a bit problematic". C129-136.

4c. Ensuring data returns

Although accessibility issues have been highlighted, and the use of paper-based questionnaires highlighted as an alternative, issues associated with this were discussed.

"We say to patients would you mind filling this in they don't generally mind doing it. If they have time to do it while they're here that's much better. The problems arise when they take it away. And if you say to them you can't come back until you have completed it then they probably won't". A, 275-279.

4d. Organisation and costs associated with PROMs

Resource use in terms of time and supply of the questionnaires was discussed as a topic of note for consideration when trying to implement PROMs into a clinical setting.

"The only practical issue really seems to be time, and that will depend on the instrument being used, or whether it's one or more than one instrument". B, 77-79.

4e. Standardisation of measurement instruments

The practicalities of data collection were discussed in terms of the non-standardised use of instruments. The role of the professional body (the Chartered Society for Physiotherapists) was suggested as a group to facilitate this process.

"I think we need some sort of system to collect and collate data about practice. I think it's a pity we don't have a universal system say through our professional body that everyone can use and which is standardised". A, 9-12.

Theme 5. Appropriateness of measurement

While the necessity of measurement in different settings was accepted, different views were expressed about the volume of data collected, and the tools used in the measurement and data collection process.

5a. Identifying instruments sensitive enough to measure change

The importance of choosing an instrument which was appropriate for the patient population and setting was discussed.

"I think the main thing with using PROMs is whether the data that is being collected is frankly too crude to have any actual value. And that could be a problem if we try and rely on that type of information which is too crude to actually show any benefit to the patient. I think my only other concern would be that too much information would be collected which wouldn't actually have any benefit to the profession or the patient". B,117-124

5b. Being aware of the necessary limitations of measurement

The tension between capturing all of the data that is relevant to a patient population with the practicalities of data collection was discussed.

"*Well you'll never capture everything. I think from a practical point of view you just have to ask what is quick and relevant. You're never going to collect data for every eventuality are you". A, 159-162.*

5c. Being clear about the purpose of measurement

While the use of PROMs was described as having useful application, it was suggested that the limitations of what could be measured, and what those measurements could demonstrate or reveal should be appreciated.

"I think the main thing to have in mind with this is to have in mind what exactly it is that you are trying to achieve. So the hardest question in my mind is what you want to get out of it, what do you want to achieve from using PROMs. That brings up the question of whether you will actually find what you want, whether there is something there or whether you will have to devise something of your own". B, 142-148.

An extensive range of views on the topic of PROMs, and their value in practice have been contributed by physiotherapy participants. These views have been summarised in the next section.

3.3.12 Evolved models and their meaning – Physiotherapists

In this set of interviews, the experience of clinicians who have worked in the NHS is evident. These views have been summarised in the model shown in Figure 3.5.

The model highlights the inter-relationship between the need to embrace new innovations *e.g.* PROMs and the practical challenges this brings. The tension of meeting professional challenges of gaining PROM completion, identifying what is useful from the analysed data, and then implementing change is highlighted. The importance of using PROMs to foster greater engagement with patients is highlighted, while at the same time valuing the feedback patients provide, and how

this can impact upon their management. In the context of delivering care in the face of growing demand and relative reductions in resources, the demands of healthcare managers and commissioners must be met. Patient burden has been discussed and the necessity to complete multiple questionnaires to demonstrate efficiency of service and the effect of practice is a professional requirement. Even within the private sector, professional challenges of data returns must be met; the introduction of new innovations has been useful to support this process. While new innovations are introduced, the capacity for measurement, and the limitations of those measurements must be considered. Patients are multifaceted individuals sometimes with multiple symptoms and complex comorbidities; single outcome measures to capture the extent of such data are noticeable for their absence.

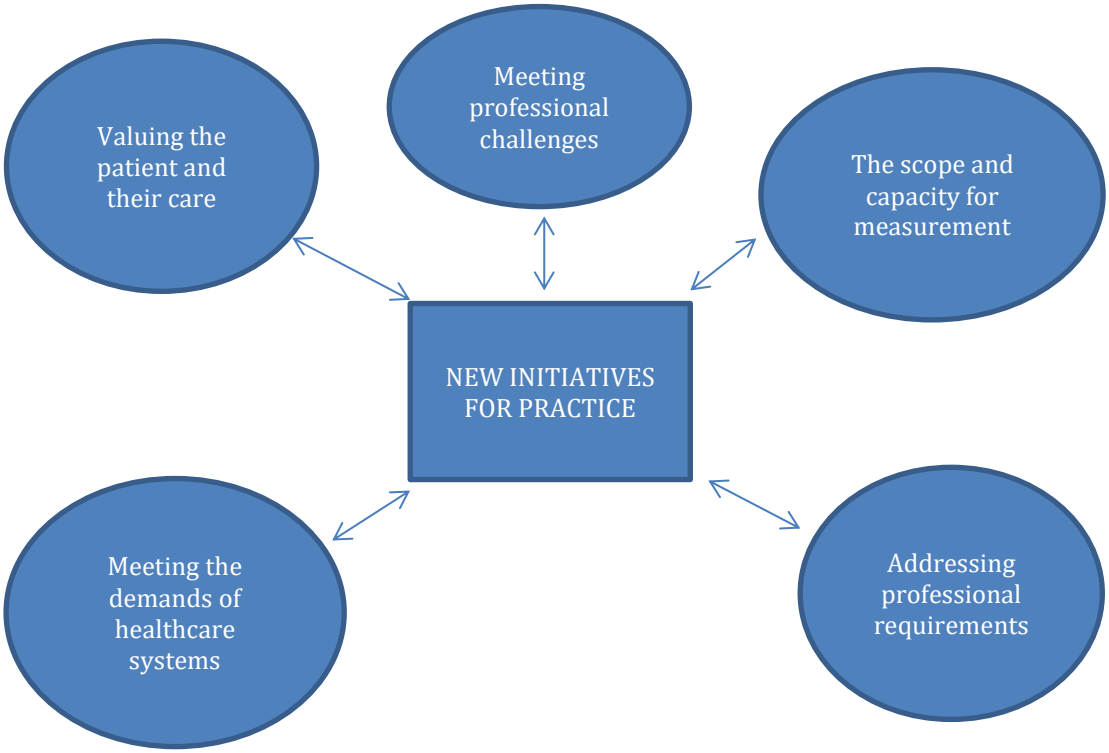


Figure 3.5 Physiotherapy model developed from physiotherapy themes and sub-themes.

Although there were some noticeable areas of agreement with the professional groups, there were also some differences. When discussing the personal development role of PROMs, many of the clinicians interviewed noted that they had approached PROMs use with slight trepidation. Although clinicians frequently

perceive that their patients respond well to treatment, this may not always be the case. As Don Herold noted (Penguin, 2001):

"Doctors think a lot of patients are cured who have simply quit in disgust".

Equally the same might be true of manual therapy clinicians, and while learning of a lack of success in patient management can make uncomfortable reading, more participants thought it was better to know and address any issues in service delivery, or technical approaches than to live in blissful ignorance. Only one participant stated explicitly that they would prefer not to find out. Participants noted that there had been occasions that they had learned about practice issues of which they were completely unaware, and the feedback at least gave them a chance to change things. After exploring the interpretation of clinicians on PROMs' use and value, it might be said that their views could reflect a continuum where increasing experience has been matched with sustained and increasing implementation in practice (Figure 3.6). This may have been due to the clinical context within which clinicians are currently working, or have worked at different points during their careers.

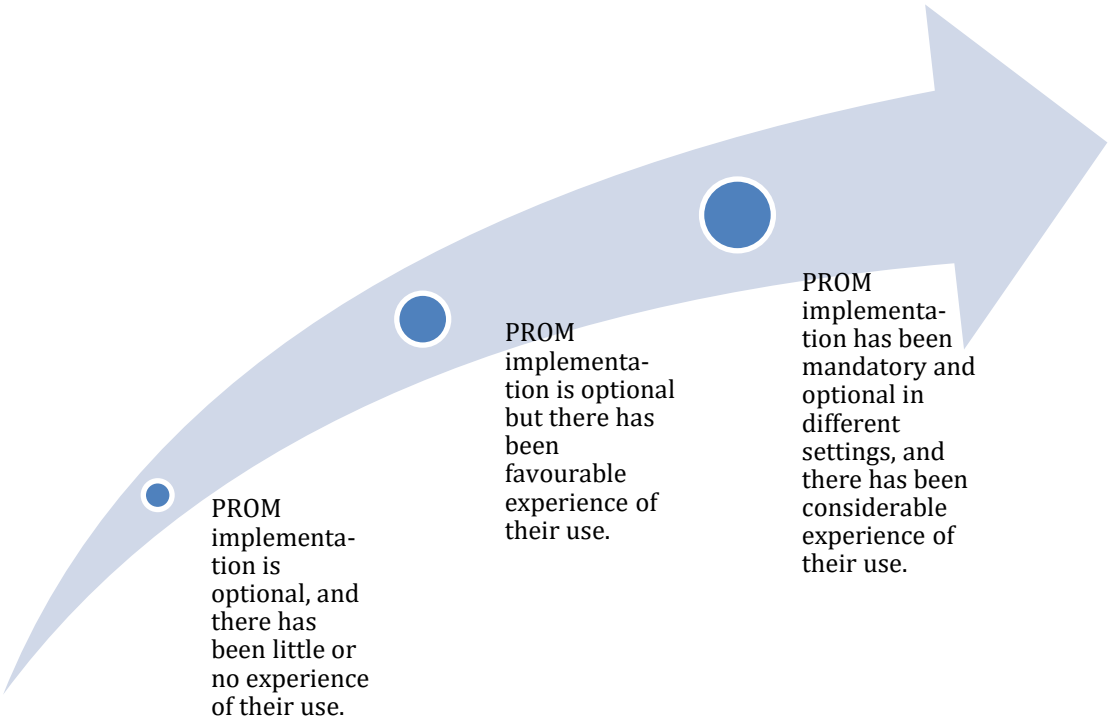


Figure 3.6 A continuum of views based on experience of using PROMs.

Although differences have been identified within the professional groups, there are some contrasts between patients' and clinicians' views. These will be explored in the Discussion section of this chapter.

3.4 Discussion

In completing this qualitative study involving three different professional groups, a range of views has emerged. These will be summarised and discussed within the context of existing research. Finally the strengths and limitations of this study, and the implications for practice and research will be discussed.

3.4.1 Main findings

When examining the participant characteristics across the professional groups, the geographic variation is good although there were no participants from Wales or Northern Ireland. Work situations were varied across all professional. Ethnic variation was quite limited with most participants describing themselves as white British. The number of years in practice was mixed with the physiotherapy participants having the largest mean number of years since qualification. The participants in the osteopathic and chiropractic participants were more varied in their years since qualification. When considering the contributions of all professional groups, key findings are identifiable

In those clinicians whose experiences and exposure to PROMs were severely limited, the concerns about the time available for completion, patients' willingness to complete, and their IT capability were noticeably more marked. The necessity to take a more pragmatic approach was discussed, emphasising the value to individual clinicians and professions as a whole while recognising also the limitations that PROMs and the data they deliver can have. Notwithstanding these issues, clinicians recognised the importance of asking patients for their feedback whether this was a mandatory requirement of their clinical setting, a demonstration of good practice, or to harness additional business opportunities.

Among clinicians who recognised the value of PROMs, there was recognition of their limits in data that could be delivered, and limitations within individuals and their ability to analyse and interpret PROM data. The need for training to be available was highlighted by some clinicians, whereas for others a system which could deliver analysed data was preferable.

Among clinicians who recognised the value of patient feedback, the potential challenges associated with this were recognised. Negative patient feedback could be recognised as part of practice life and an opportunity to learn from the patient encounter; for others being informed that their practice was not as good as they thought was quite discomforting and unwelcome.

The findings from the three professional groups in this study have been valuable. These findings will be placed within the context of previous research in section 3.4.2.

3.4.2 Previous research

The concept that PROMs should reflect the "uniqueness" of some interventions was mentioned by one professional group; however the lack of capacity for any PROM to measure all of the symptoms with which patients present was noted by all (Valier *et al.,* 2014). Some participants noted that fundamental issues may arise due to the manner in which patients with musculoskeletal symptoms are characterised. Clinicians look for key symptoms *e.g.* pain or disability because they can be an easy basis on which to discuss limitations of movement, or experiences that cause patients enough distress to seek treatment in the first place. The literature tells us that patients present with single or multiple comorbidities which may or may not have an emphasis on the aetiology of their symptoms or subsequent recovery (Nelson *et al.,* 1983; Valderas *et al.,* 2008a). A pragmatic approach focussing on the issues which are most apparent is often borne out of necessity in clinical management, and the same could extend to the measurement of outcome (Marshall *et al.,* 2006). There is, however, an unrecognised tension in clinicians desiring patients complete a PROM which reflects all of the different symptoms which may

be important to them, while at the same time highlighting the burden on patients in being asked to complete one long or several questionnaires.

The capacity of PROMs to act as effective measures of change was discussed. There are thousands of PROMs available and although some will have been created according to development guidelines, this will not be the case for all of them (Nelson *et al.,* 2015). I will explore the features around measurement properties in greater detail in Chapter 4, but it should be recognised that many of the PROMs in use in mainstream clinical practice in primary, secondary, and tertiary care settings have been developed in a robust manner, and it is important to select PROMs which have been developed for a particular setting, involved patients, and have been well-tested in similar settings in which they will be employed.

For some clinicians, using PROMs is a mandatory part of their professional life to satisfy the requirements of NHS managers, and commissioners of their service (Gurry *et al.,* 2004). Other clinicians noted that they saw PROMs as the means of seizing opportunities available for care delivery resulting from political changes. The use of PROMs simply as part of good practice, as a means of allowing patients to have their say, or engaging patients in their treatment process were the motivations for other clinicians (Barry *et al.,* 2012). None of these characteristics are mutually exclusive. While viewing PROMs and their potential to deliver valuable clinical and practice information was important to some clinicians, others approached their use in full recognition that they could encroach upon patient time if not implemented in a manner to specifically avoid this. Others recognised that they possessed the "zeal of the convert" since starting to use PROMs which could make them less receptive to the practical issues and shortcomings inherent in any measurement approach applied to a diverse population.

Notwithstanding the shortcomings of the actual PROMs, the challenges for other clinicians were in using them in settings where multiple languages may be spoken, and low literacy levels (Valier *et al.,* 2014). This was most notable in the clinicians who had used them in NHS settings where a more diverse population attended for treatment, and whose attendance was not based upon ability to pay at the point of

care. The issues of IT access and social exclusion was discussed in Chapter 2, but even in those practices where IT access had been available in the form of Tablet computers, there were still difficulties associated with this in the form of basic screen dexterity, and the issues of failing sight and lack of glasses experienced by many age groups. Nonetheless, despite the challenges of such practical issues, many clinicians felt that there would be an increasing expectation from patients for feedback to be sought. This is an increasingly common practice within the NHS, and from private medical/health insurers, and it might be viewed as an anomaly if private practices were not acting in a similar manner. For some clinicians, however, they felt they had a lack of knowledge about how to use, analyse, and interpret the findings from PROMs and this underpinned some of their reluctance to use them (McAuley *et al.,* 2014; Valier *et al.,* 2014).

Finally for some of the clinicians, they recognised a frustration that some sections of their profession would not engage with PROMs, or data collection of any kind as it was a challenge to their philosophical stance, or because they just lacked the motivation. This was a position that seemed unrelated to years since qualification since it was noted by more recent post-registration clinicians as well as those of longer duration. Regulatory requirements are the only factors likely to change this stance, but for those who had used PROMs they felt that clinicians not using PROMs were failing to benefit from the personal development that PROMs could offer.

Comparing patients' and clinicians' views

Although differences emerged concerning PROMs' use in the areas regarded as more and less important amongst clinicians, there were also clear differences in the views between patients and clinicians. Themes are mapped according to areas which were regarded as limiting factors/areas of concern, neutral factors, positive factors, and areas where there were mixed views. This is shown in Table 3.7.

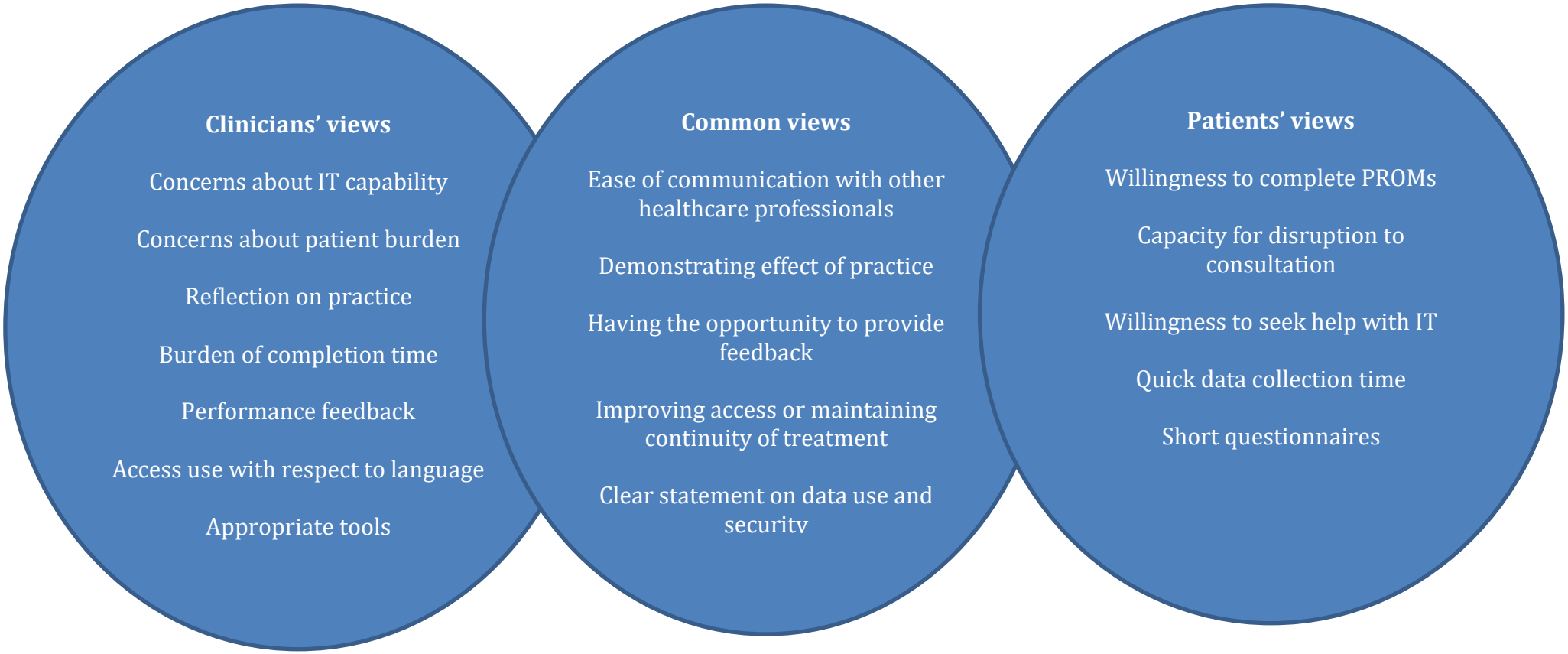
Table 3.7 A comparison of themes between each of the professional groups and patient participants

Limiting factors/areas of concern	Neutral	Positive factors	Mixed views in favour and resistant to PROMs' use
Relevance and appropriateness to patients (including simplicity	Variety of formats	Information about practices (types of problem treated *etc*)	Comparative data
Use of data	Technological issues in completion modes	Measuring effectiveness of treatment	Using electronic resources
Confidentiality in technology	Clinical vs. research purposes	Consideration/reflection on other issues	Demonstrating progress/effectiveness
Data protection	Feedback within consultation	Widening patient access/choice	Marketing
Clear statement of purpose of data	Information overload	Information for other healthcare professionals	Wider clinical issues
Value of PROM data	Potential to measure change	Variety of dissemination formats	External support
Motivation	Patient follow up	Choice of participation	Management appraisal
Patient burden	Independent evaluation	Dissemination	Supporting research
Lack of confidence	Reflection/discussion	Business development	Identifying practical issues
Security of data		Communication with other healthcare professionals	Keeping up to date with new initiatives
Analyses of data		Professionalism	Demonstrating efficiency of practice
Cost/resources		Identifying practical issues	Patient engagement in the treatment process
Implementation into practice		Identifying clinical issues	Identifying what is important to the patient experience
Therapeutic relationships		Evaluating the service delivered	Standardisation of measurement instruments
Training			
Language consistency/comprehension			
Philosophical and personal challenges			
Use of practice resources			
Added bureaucracy for staff and patients			
Dealing with the demands of external agencies			
Challenges to existing skills			
Different levels of literacy and language			
Avoiding negative impact on the consultation time			
Being aware of the limitations of measurement			
Being clear about the purposes of measurement			

Patients' views alone Clinicians' views alone Views shared by patients and clinicians

A range of issues can be identified when considering the different views of clinicians and patients. There are areas that are common to all groups; equally there were views that are unique to either the clinicians or patients. These are summarised in Figure 3.7. It is clear in Figure 3.7 that there are some important areas of common agreement between patients and clinicians concerning the use of PROMs in practice, there are some notable areas which are unique to each group.

Figure 3.7 A summary of views of commonality and separateness for patients and clinicians when discussing PROMs



While the clinicians expressed concern about IT capability, notably with older patients in mind, patients suggested that they were happy to complete the PROMs and if they needed help to access then they would locate it among friends and family. Equally, those patients who did not own a computer or smartphone were motivated enough to complete the PROM with family who did have IT access. Patients' main comment about completing the PROM was that they would be reluctant to do it during appointments as it would reduce clinician contact time, and could potentially disrupt the flow of the consultation. Patients did make it clear that they would be willing to spend a short amount of time completing the PROMs (ranging from 5-20 minutes), and they were not especially concerned with getting information back unless it was in the form of monitoring information related to their own problem/symptoms.

Areas where patients and clinicians agreed were about the potential value of the data to external stakeholders. For patients who attended funded services in the NHS, data were about maintaining access to that service; while for others it was useful material to support applications allowing access to publicly-funded care. The use of the data for communicating with other healthcare professionals to describe treatment and what it could potentially offer was regarded as valuable also. All groups were quite clear concerning the use of data and security. A clear statement on the use of data should be supplied to clinicians and patients, and they should have confidence in the anonymity and security of the data capture facility.

3.4.3 Strengths and limitations

Despite the focus on the use of PROMs in publicly-funded healthcare, and by private healthcare insurers, there is a paucity of literature investigating clinicians' views on their experiences of their use. The greatest focus of the literature to date has been on the perceived barriers and facilitators to PROMs' use, and this has a greater contribution from international researchers. This might be explained by the fact that Organisation for Economic Cooperation and Development (OECD) data shows that spending on healthcare in OECD countries (excluding the extensive spending of the USA on healthcare) represents 9.1% of Gross Domestic Product (GDP). In 2013, levels of spending on public and private healthcare in the

UK placed us 13th out of the 15 original European Union (EU) countries (Appleby, 2016). The need to research the views of healthcare professionals in the UK is important, and this is one of the strengths of this work. It has contributed to knowledge by interviewing clinicians from three separate disciplines. This can help to inform implementation strategies to make the use of PROMs in clinical practice easier within a busy practice setting, and produce more valuable output for the benefits of patients and clinicians alike.

Inevitably, there will be limitations to any research. The sample obtained for the interviews contained greater numbers of osteopaths and chiropractors than physiotherapists. This was frustrating since there were reported to be 7,300 physiotherapists working exclusively outside of the NHS in 2010 compared with 5,193 osteopaths, and 3,109 chiropractors (Beddow, 2010; GOsC, 2016a; GCC, 2015). It would have been helpful to have larger representation from solely private physiotherapists with a greater age range included.

The decision to conduct the interviews by telephone was a practical issue. Participants were located at geographically diverse locations making travel difficult while trying to complete this aspect of the study within a defined academic programme. Although it has been suggested that telephone interviews fail to allow capture of nuances in facial expression and body language that can add to the richness of qualitative data (Garbett and McCormack, 2001; McCoyd and Kerson, 2006). The importance of face-to-face as opposed to telephone interviews is much debated in the literature (Novick, 2008; Sturges and Hanrahan, 2004). There is a lack of comparison studies, and there are clear views on the value of varying the approaches for different types of data, subject areas, and populations (Groves, 1990). Among the small number of head-to head studies, very little difference in the data has been revealed (Tausig and Freeman, 1988; Sturges and Hanrahan, 2004). The interview approach must be employed with consideration to widening access for both participants and researchers, comfort of the participant, researcher safety, and cost (Aday, 1996; Chapple, 1999; Carr and Worth, 2001; Bernard, 2002; Novick, 2008). The ability to complete some face-to-face qualitative work through

the focus groups has aimed to mitigate any perceived shortcomings in using telephone interviews with some participants.

In contrast to criticisms about interviews, focus groups are recognised to have potential issues where a type of dynamic can arise when certain group members are more assertive or compelling than others in making their views known. This can influence other group members and introduce a form of "group bias" (McGee, 1999). I was prepared for dealing with this issue, although many of the osteopathic participants knew each other already through their involvement in the NCOR research hubs and regional society meetings, and this did not occur in practice.

While identifying clinicians' views about the development of a data capture system has been instructive there are strengths and limitations to interviewing of osteopaths by another osteopath. Most of the osteopaths (90%) participating in the focus groups and interviews knew me already so I only introduced myself as a practising osteopath in some of the focus groups where participants were unfamiliar to me (Leeds and Bristol). In the interviews, my role as an osteopath was made clear in the recruitment messages to potential participants. An interviewer who was not an osteopath may have explored more issues about which osteopaths held concern, and this aspect of the study has required considerable reflection from the perspective of my own performance, and strenuously avoiding leading questions, or failing to listen to dissenting views about the whole project. The alternative would have been two interviewers from different professions which may have undermined the consistency of the approach.

Trustworthiness and credibility of findings

The concept of reliability (Ritchie and Lewis, 2003) or trustworthiness (Guba, 1981) in qualitative research was discussed in Chapter 2. Guba proposed four criteria should be considered in pursuit of trustworthiness which includes credibility - the believability of the findings - (Guba, 1981). Methodological concerns which relate to trustworthiness and credibility within qualitative research include, for example, the sample coverage, the environment for capture of

the data, the identification or labelling of the data, the manner of interpretation of the data, and the display of the data and how this remains true to the original data (Guba, 1981; Lincoln and Guba, 1985; Bowen, 2009; Holloway and Wheeler, 2015). These concepts are as they relate to this qualitative study are described in Table 3.8

While this qualitative study has explored a range of helpful issues, there are areas which could be developed further. These will be discussed in the section 3.4.4.

3.4.4 Implications for the app

The views of clinicians in this qualitative study have been illuminating. Although their perspectives echo the findings of existing research, they have identified new areas of importance. They have highlighted a range of issues, as have patients, which need consideration when designing and implementing an electronic PROM data capture system. Participants from all professions have highlighted the importance of the methodological robustness of PROM data collection, and consistency of outcome reporting. Underpinning these issues is the need to be aware of bias include from incomplete PROM reporting (Higgins *et al.,* 2011).

To ensure successful implementation there needs to be a shared vision agreed between patients, healthcare professionals, purchasers, *etc* concerning the use and application of data to maintain existing trust and promote confidence in the data collection process (van der Wees *et al.,* 2014). The role of opinion leaders or champions to advocate the added value for clinicians will be vital in this process, and the qualitative work has suggested how this could be undertaken to resonate with osteopaths (van der Wees *et al.,* 2014).

When beginning a new project, enthusiasm is at its height, and it is important to be realistic about what it can deliver, and also the need to revisit the original aims of the project to ensure they are still relevant within changing healthcare settings and political context. It was important for this new initiative to start with a small pilot

Table 3.8. Strategies used to enhance the trustworthiness of the study (Guba, 1981; Lincoln and Guba, 1985; Holloway and Wheeler, 2015).

	Description	Strategies
Credibility	Confidence that the research has provided an accurate determination of the meaning behind the data thereby reflecting the clinicians' experiences of using PROMs, and identifying any concerns or opportunities associated with their use.	**Immersion in the data.** Time was spent initially engaging with all of the data emerging from each of the professional groups to become immersed in the richness of the finings; **Member checking**; A sample of participating clinicians were invited to read through the verbatim transcripts of their interviews to confirm their accuracy, and have the opportunity to request removal of any comments with which they were later uncomfortable. Peer debriefing: discussion of the findings with individuals (researchers and non-researchers not involved in the study to provide feedback on the coding and analysis of findings.
Confirmability and dependability	Whether the findings from the study provide a dependable and realistic presentation and interpretation of the views expressed by clinicians from three professional groups.	Audit trail. The importance of maintaining field notes, memos, and other recordings of the research process either in visual or written forms. Maintaining notes on the stages undertaken and decisions made during the analysis process whether singly

		or in collaboration with other researchers. The researcher was able to remain reflexive towards her involvement in the interview process, and the impact this could have on questioning and interpretation of the findings. Participants were reminded that their views were important and there were no right or wrong answers. Participants were asked to be as forthright as possible with their views on the potential use of PROMs in clinical settings. They were reassured that their data would be anonymised and at no time would they be identified in any form of publications of the findings of the research.
Transferability	The extent to which the ideas generated by this research may be applied to other clinical settings and patient populations.	Writing accurate accounts based on the rich data provided by the clinician, and providing information that could be applicable to a range of different research settings.

involving willing clinicians to overcome initial resistance (van der Wees *et al.,* 2014), making it clear that the project was a learning process which will be reflexive to changing circumstances.

3.4.5 Future research

The interviews and focus groups have indicated that it is important to teach clinicians the value of PROMs, how to interpret the data, and change scores irrespective of whether they have calculated them as individuals or it has been accomplished by a third party (van der Wees *et al.,* 2014). The role of the PROM in clinical care has been shown to foster communication, and it is important for clinicians to discuss the use of PROMs and how the patient felt when considering some of the questions (Velicova *et al*., 2004; Swinkels *et al.,* 2011; van der Wees *et al.,* 2014). This dialogue can be instructive in clinical management even without access to changing PROM scores within the consultation.

The effect of PROMs' use on clinical practice, and outcomes of care is an area for considerable future research. Future studies could examine through qualitative methods whether use of the app has increased clinicians' knowledge and changed their practice. Whether this changing practice translates into improved outcomes could be investigated with a quantitative approach involving PROMs' users and clinicians not using PROMs.

While there is an assumption that clinicians' use of PROMs will continue uninterrupted over time, it will be important to explore any reasons for lack of engagement either at the start of the process or at some point after initial use of PROMs has been undertaken. The use of PROMs should be reviewed continually and support be responsive to the changing needs of clinicians.

3.4.6 Conclusions

The collection of PROM data can directly inform clinicians and future patients regarding the effects of an intervention, or may influence clinical practice when large volumes of data concerning particular interventions are analysed (Efficace *et al.,* 2014). It is an important aspect to clinical care, and is increasing in its

implementation across all healthcare delivery. To engage osteopaths it was vital to identify their views on such an initiative and identify some of the practical challenges they perceived in trying to implement such a process. This qualitative study, in consideration with the views contributed by patients (described in Chapter 2), has been fundamental in the process of developing a PROMs data capture facility for osteopaths.

4.

Systematic review of the measurement properties of the Bournemouth Questionnaire

4.1 Introduction

Three important aspects of healthcare and patient management are diagnosis, evaluation of treatment and management approach, and the prediction of the future course of a patient's symptoms/disease. To foster such measurement in healthcare, several disease-specific Patient Reported Outcome Measures (PROMs) have been developed to measure pain and disability in patients with musculoskeletal pain. The science of such measurement has been termed clinimetrics (Feinstein, 1987). There has been much debate concerning the term "clinimetrics" and whether it holds an arbitrary distinction from psychometrics (de Vet *et al.,* 2003; Streiner, 2003; Emmelkamp, 2004; Fava *et al.,* 2012). Healthcare invoves several different forms of metrics and there is some agreement that the term clinimetric should be reserved for multidimensional health measurement scales and indices (Feinstein, 1987; Fayers and Hand, 2002). In order to ensure

clarity throughout this chapter and thesis, the clinimetric definitions used are described in greater detail in the next section.

Notwithstanding the importance of good measurement properties of individual PROMs, they require adaptation also for use in other languages. Translation alone does not guarantee a duplication of the original measurement properties of a PROM due to the influence of the cultural context in which it might be used, and its effect on performance (Beaton *et al.,* 2000; Wang *et al*., 2006). This can lead to inconsistencies in the performance of a PROM when used in different settings (Menezes da Costa *et al.,* 2009). In recognition of this issue, the COSMIN group have included translation and cross-cultural validity as a feature in their assessment tool of the measurement properties of PROMs.

The taxonomy described by the COSMIN group is summarised visually at http://www.cosmin.nl/images/upload/files/COSMIN%20taxonomy.pdf. It is summarised diagrammatically in Figure 1.7 in Chapter 1.

4.1.1 Clinimetric terminology

Concept.

In PROMs a concept represents the specific measurement goal (de Vet *et al.,* 2011).

Content validity

This examines the extent to which the concepts of interest to a particular population or condition are represented by the items in a questionnaire. To be able to evaluate a questionnaire, there should be agreement on various different aspects which are implicit in the questionnaire development. These include:

i. The measurement aim of the questionnaire being assessed;
ii. The target population the questionnaire is evaluating;
iii. The concepts the questionnaire is intended to measure and their suitability for their intended purpose;
iv. The methods chosen for item selection and item reduction;
v. Whether a pilot study has been undertaken to examine readability and comprehension;

vi. The interpretability of the items and whether a reading age in excess of 12 years is required to understand those items (de Vet *et al.,* 2011).

Construct validity

This refers to the extent to which the scores on a particular instrument relate to other measures/instruments in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being evaluated. This should be assessed by:

i. Testing predefined hypotheses which are as specific as possible *e.g.* expected correlations between measures;
ii. Testing predefined hypotheses *e.g.* expected differences between "known" groups;
iii. Identifying whether a positive rating can be said to exist;
iv. Identifying a positive rating when hypotheses are stated in advance, and when at least 75% of the results are in correspondence with the stated hypotheses in subgroups of at least 50 patients (de Vet *et al.,* 2011).

Criterion validity

This refers to the extent to which scores on a particular instrument are an adequate reflection of a gold standard. Rating for criterion validity will be regarded as positive if:

i. A convincing argument is proposed that the "gold" standard represents the best available standard;
ii. The correlation with the proposed "gold" standard is at least 0.70 (de Vet *et al.,* 2011).

Cross-cultural validity.

This is the degree to which a translated or culturally adapted version of a measurement instrument reflects the performance of the items in the measurement instrument from which it was originally developed (de Vet *et al.,* 2011).

Detection of change.

This is the ability of an instrument to identify important changes within a

population and concept of interest by examining the differences in scores of a measurement instrument. The ability to detect change occurs over time and can be examined at either an individual or population level (de Vet *et al.,* 2011).

Face validity.

This is the degree to which the items contained within a measurement instrument reflect the particular construct being measured (de Vet *et al.,* 2011).

Internal consistency

This is a measure of the extent to which items within a questionnaire or within a particular subscale are homogenous /correlated. This identifies whether they are measuring the same concept. Internally consistent scales are achieved through:

i. Good definitions of the constructs;
ii. Good items;
iii. Ensuring the correct level of measurement difficulty as typified by the Goldilocks zone (Raw *et al*., 2015);
iv. Conducting principal component analysis or exploratory factor analysis;
v. Completing with confirmatory factor analysis (de Vet *et al*., 2011).

Interpretability.

This is the capacity to be able to assign clinical or commonly understood meaning to a patient's response. This will be reflected in a quantitative response or the change in score for a particular instrument (de Vet *et al.,* 2011).

Item.

This relates to an individual question or statement to which the patient will be asked to respond. An item will relate to a particular concept that is being measured (de Vet *et al.,* 2011).

Measurement error.

This is the systematic and random error in a patient's score. This error cannot

be attributed to true changes in a patient being reflected in the construct of interest being measured (de Vet *et al.,* 2011).

Measurement properties.

This is a collection of all of the attributes within a measurement instrument. These can include the reliability of the measure, different aspects of validity, and the capacity of the instrument to detect change. The attributes are instrument-specific and will be informed by the context within which the instrument is used *e.g.* the population, clinical setting, and purpose for using the measurement instrument (de Vet *et al.,* 2011).

Minimally Important Change (MIC).

This is the threshold of change that can be considered minimally important in questionnaire/PROM data when considering individual patient data (de Vet *et al.,* 2011).

Minimally Important Difference (MID).

This is the difference in aggregate questionnaire/PROM scores that can be considered minimally important when considering data at the population level (de Vet *et al.,* 2011).

Proxy-reported outcome.

This is a measurement about a patient provided by someone other than the patient *e.g.* a carer or parent (Li *et al.,* 2015). This is different to an observer report which occurs when a clinician reports an observation concerning a patient and may also interpret that observation. They are also known as parent proxy outcomes. In some disciplines, particular domains can be measured as a proxy for another *e.g.* in the field of schizophrenia research, cognitive outcomes are assessed to act as predictors for functional outcomes (Green *et al*., 2004).

Recall period.

This is the period of time about which participants are asked to reflect when considering their responses to a measurement instrument. A recall period in

measurement assessment tends to be anything from the past 24 hours to the past few weeks; this is largely determined by the concept of interest being investigated (de Vet *et al.,* 2011).

Reliability.

This is the degree to which a measurement instrument is free from measurement error. This can be demonstrated by evaluating the extent to which scores remain the same when patients whose health status has not changed complete measurement instruments at different points in time. The assessment of a measurement instrument over time in this way is known as test-retest reliability. When a measurement instrument is examined to see if the score remains the same when used by different people on the same occasion, this is termed inter-rater reliability. If used by the same person on different occasions this is termed intra-rater reliability (de Vet *et al.,* 2011).

Responsiveness.

This is the ability of a measurement instrument to detect important change over a period of time in the particular construct being measured (Terwee *et al*., 2007).

Scale.

This is a quantitative or qualitative system for measuring responses to an item of interest. Participants' responses to a scale, whether captured by numerical or verbal means, can be used to calculate a score. A range of scales exist depending on the item of interest being evaluated (de Vet *et al.,* 2011).

Score.

This is a value obtained based on participants' responses to a measurement scale (de Vet *et al.,* 2011).

Sign.

This is an objective indicator of a disease, health condition, or the effect of a particular therapeutic intervention. It is usually observed, measured, or

interpreted by a clinician while it may also be apparent to a patient (de Vet *et al.,* 2011).

Structural validity.

This is the degree to which the scores obtained by a measurement instrument reflect the dimensionality of the particular construct being measured (de Vet *et al.,* 2011).

Symptoms.

These are a collection of subjective items which can be noticed and known only by a patient. Symptoms can relate to a particular health condition or the effects of a particular intervention as part of patient management (de Vet *et al.,* 2011).

Test-retest reliability.

This is an evaluation of the extent to which scores remain the same when patients whose health status has not changed complete measurement instruments at different points in time (de Vet *et al.,* 2011).

Treatment benefit. This is the effect of a therapeutic intervention on a patient who has reported a particular set of signs and/or symptoms. Some of these definitions are summarised in tabular form at http://www.cosmin.nl/images/upload/files/Tabel%20met%20definities-new.pdf.

4.1.2 Organisations involved in PROMs

While different research groups and individual researchers are involved in discussions concerning PROMs terminology, they are also involved in recommendations concerning which PROMs to use in clinical trials or day-to-day data collection. A summary of these recommendations is provided in Table 4.1.

Table 4.1. A summary of recommendations concerning PROM use from special interest groups/organisations

Group	Recommended Outcome	Recommended Measurements/PROMs	Recommended time line
International Consortium for health Outcomes Measurement (ICHOM)(Clement *et al.*, 2015)	Health Related Quality of Life (HRQoL)	EQ5D-3LEQ-VAS	Baseline; index event(s); 6 months; 1 year; 2 years.
	Pain	Numerical Pain Rating Scale (NPRS)	Baseline; index event(s); 6 months; 1 year; 2 years.
	Disability	Oswestry Disability Index (ODI)	Baseline; index event(s); 6 months; 1 year; 2 years.
	Work status	What is your current status?	6 months; 1 year; 2 years.
	Analgesic use	Do you take narcotic medication or tablets for your back pain? Do you take non-narcotic medication or tablets for your back pain?	Baseline; index event(s); 6 months; 1 year; 2 years.
Core Outcome Measures in Effectiveness Trials (COMET)Chiarotto *et al.,* 2015	Physical functioning; Pain intensity; Health-related quality of life; Number of deaths.	Not available Not available Not available Not available	Not available Not available Not available Not available

Multinational Musculoskeletal Inception CohortStudy (MMICS) Statement			None recommended
Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT)	Pain; Physical functioning; Emotional functioning; Participant ratings of improvement and satisfaction with treatment; Symptoms and adverse events; Participant disposition (e.g. adherence to the treatment regimen and reasons for premature withdrawal from the trial).		None recommended
Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT)*Boers et al.,* 2014	Pain Tender joints Swollen joints Function Patient's opinion Clinician's opinion X-rays		None recommended
World Health Organisation (WHO)(WHO Scientific Group, 2003)	WHO Quality of Life(WHOQOL); WHO Disability Assessment Schedule	WHO Quality of Life(WHOQOL); WHO Disability SF-36 (Generic) Oswestry (Specific) Roland Morris Kohlmann–Raspe	None recommended
Representatives from BUPA heathcare insurers (Swarbrick, 2013)	Pain and functional disability Patient satisfaction	ODI; RMDQ; Bournemouth Questionnaire	None recommended

Any Qualified Provider (AQP)/Clinical Commissioning Groups (CCGs)	Pain and function Quality of life Patient satisfaction	Lack of consistent measures recommended. Reported measures include BQ and EQ5D	
Deyo *et al.,* 1998	Functional disability	RMDQ ODI	None recommended
Bombardier *et al.,* 2000	Pain Functional disability Quality of life Patient satisfaction	Roland-Morris Disability Questionnaire Oswestry Disability Index;SF-36;Patient satisfaction	None recommended

While several different groups have given their recommendations on which measurements to use, it is worthwhile to reflect on what we know about some of these measurements. In 2004, Müller *et al.* undertook a review of measures used in spine surgery for low back pain and identified 82 separate measures and scales. In contrast, Chapman *et al.,* describe the most frequently cited outcome measures in the measurement of low back pain (Chapman *et al.,* 2011). The findings from both studies are described in Table 4.2.

Table 4.2. A summary of the most frequently cited PROMs used in the measurement of low back pain and spine surgery.

Muller *et al.,* 2004	Chapman *et al.,* 2011
Oswestry Disability Index (ODI)	Oswestry Disability Index (ODI)
Roland Morris Disability Questionnaire (RMDQ)	Roland Morris Disability Questionnaire (RMDQ)
Low Back Outcome Score (LBOS)	Quebec Back Pain Disability Scale (QBPDS)
Quebec Back Pain Disability Scale (QBPDS)	Functional Capacity Evaluation (FCE)
Million Visual Analogue Scale (MVAS)	Low Back Pain Rating Scale (LBPRS)
Aberdeen Low Back Disability Scale (ALBDS)	Schober's Test
North American Spine Surgeons Lumbar Spine Outcome Assessment Instrument (NASS LSO)	Patient Specific Functional Scale (PSFS)
Low Back Pain Rating Scale (LBPRS)	Trunk flexion
Waddell Disability Index (WDI)	Physical strength

When considering content for a PROM app for low back pain, a range of issues needed to be considered including the frequency of use of different measures, their measurement properties but also their development. As part of this review, the measurement properties of the ODI and RMDQ were reviewed to assess whether they would be suitable PROMs to include in an app for the specific use of osteopaths in private practice. However, based on the findings of the qualitative work described in earlier chapters, the Bournemouth Questionnaire is the main PROM of interest.

4.1.3 Aims and objectives

The aim of this review is to systematically examine and appraise existing research concerning the development and measurement properties of the Bournemouth Questionnaire (BQ). The systematic review will allow a comparison to be made with two other PROMs to identify which will be more suitable to include in a PROM app developed as part of this PhD.

The objectives were to undertake a systematic review of the measurement properties of the Bournemouth Questionnaire and assess:

i. the validity of the Bournemouth Questionnaire (low back and neck versions);
ii. the reliability of the Bournemouth Questionnaire (low back and neck versions);
iii. the responsiveness of the Bournemouth Questionnaire, and values associated with responsiveness;
iv. the validity of translated versions of the BQ neck and low back versions, and cross-cultural validity of the translated versions;
v. the suitability of the BQ for use in a PROM app when compared with the Oswestry Disability Index, and the Roland Morris Disability Questionnaire.

4.2 The Systematic Review

"*If, as is sometimes supposed, science consisted of nothing but the laborious accumulation of facts, it would soon come to a standstill, crushed, as it were, under its own weight.*" – *(Lord Rayleigh, 1884)*

This statement was made in 1884 by Lord Rayleigh, Professor of Physics at Cambridge University. It is a sentiment that could equally be echoed today by clinicians attempting to keep abreast of research output to allow the delivery of evidence-informed practice. Bibliometric analysis has reviewed research output over many decades. Output has been estimated to be increasing at 9-

10% per year (de Solla Price, 1970; Chalmers *et al*., 2002; Bornmann and Mutz, 2015) equating to a doubling of global scientific output every nine years (Van Noorden, 2014). One attempt to synthesise this volume of data to manageable quantities has been the introduction of the systematic review. While research synthesis has been developed in many scientific disciplines since the early 20th Century (Pearson, 1904), the introduction of standards to promote scientific rigour has been a more recent innovation (Mulrow, 1987 ; Oxman and Guyatt, 1988).

Suboptimal reporting of systematic reviews and meta-analyses having been recognised, an international group developed guidance in 1996 termed the QUORUM statement (Quality Of Reporting Of Meta-analyses). In recognition of the desire to include both systematic reviews and meta-analyses in reporting guidelines, QUOROM was revised and renamed PRISMA (Preferred Reported Items for Systematic reviews and Meta-Analyses (Moher *et al.,* 1999). In addition to revising the standards of reporting, attention focussed also on the variations in terminology concerning the measurement properties of PROMs, and the manner in which their quality should be appraised.

This chapter will focus on appraising the measurement properties of the BQ, will evaluate those measurement properties according to consensus guidelines, and report the findings using the PRISMA guidelines.

In Chapter 1, I described the range of outcome measures used in clinical trials, and key PROMs employed when assessing low back pain. While the use of PROMs is now expected in clinical trials, the quality of the different PROMs available should be considered carefully prior to their inclusion. The utility of PROMs in day-to-day clinical practice is also being sought increasingly as PROMs become part of the evaluation of routine clinical care. However, when using PROMs in day-to-day practice due consideration is required concerning:

- The setting in which the PROM was developed;
- Its reliability;
- Its validity for the population and setting for which it will be used;

- Its ability to detect change within the population and setting where it will be used;
- Its cultural appropriateness in a setting with multiple ethnic groups.

Reviews of the measurement properties of the ODI and RMDQ have been undertaken already (Leclaire *et al.,* 1997; Davidson and Keating, 2002; Smeets *et al.,* 2011). They have been compared also with other PROMs. However, the BQ, developed for a primary care manual therapy setting, has not been reviewed with respect to its measurement properties. As the BQ was one of the small number of PROMs chosen for potential inclusion in the PROM app, it was necessary to focus on the BQ for this systematic review. In addition, the BQ was included in a small selection of PROMs being considered by UK insurers and Clinical Recommendations of UK Clinical Commissioning Groups (personal communication with Greg Swarbrick, BUPA).

A range of quality appraisal tools has been developed to support systematic reviews examining different methodological approaches and diagnostic procedures. Several attempts had been made to develop a checklist to examine measurement properties of PROMs before the COSMIN group published their tool (Valderas *et al.,* 2008; Mokkink *et al.,* 2010a). The COSMIN checklist contains twelve boxes under the headings:

- Internal consistency;
- Reliability;
- Measurement error;
- Content validity;
- Structural validity;
- Hypotheses testing;
- Cross-cultural validity;
- Criterion validity;
- Responsiveness.

Ten boxes can be used to assess whether a study meets the standards required for good methodological quality. Nine of the boxes contain aspects of measurement

properties. Two additional boxes are included that contain general requirements for articles in which Item Response Theory (IRT) is used, and requirements for generalisability of the results. The tool contains 100 questions in total (Mokkink *et al.,* 2010b). The measurement properties of the BQ evaluated using the COSMIN tool can be found later in this chapter (section 4.4.2). First I will explore the basic characteristics of the BQ.

4.2.1 The Bournemouth Questionnaire

The Bournemouth Questionnaire (BQ) is a seven-item questionnaire containing five categories:

General activity	(Housework, washing, dressing, walking, climbing stairs, getting in/out of bed/chair);
Anxiety/depression	Anxious (tense, uptight, irritable, difficulty in concentrating/relaxing); depression (down-in-the-dumps, sad, in low spirits, pessimistic, unhappy);
Work	(Both inside and outside the home);
Pain	Rating of pain; Ability to control (reduce/help) own pain;
Leisure	Participation in recreational, social, and family activities.

Although existing back-specific questionnaires were available, they were regarded as long, cumbersome, and requiring both time and expertise to administer and interpret (Bolton and Breen, 1999). The BQ development recognised the necessity to measure the multidimensional aspects of back pain experience within a population of ambulatory back pain patients in a chiropractic outpatient clinic. The instrument was developed based on methodological frameworks of Kirschner and Guyatt, and Streiner and Norman, and intended to be quick and easy to use, based on the conceptual model of back pain, acceptable to patients in content and avoiding being burdensome, valid, reliable, and responsive to change (Kirschner

and Guyatt, 1985; Streiner and Norman, 1995; Bolton and Breen, 1999). Preliminary testing of the instrument was undertaken to assess face validity (Phase 1), homogeneity and reliability (Phase 2), validity and responsiveness (Phase 3) (Bolton and Breen, 1999). In Phase 3, the BQ was validated against the Chronic Pain Grade (CPG) questionnaire, the Revised Oswestry Disability Questionnaire (RODQ), the Distress and Risk Assessment Method (DRAM), the Pain Locus of Control (PLC), the Fear Avoidance Beliefs Questionnaire (FABQ), Zung self-rating depression scale, and Modified Somatic Pain Questionnaire (MSPQ) (Smith *et al.,* 1997; Hudson-Cook *et al.,* 1989; Main et al, 1992; Toomey *et al*., 1995; Waddell *et al.,* 1993; Zung, 1965; Main, 1983).

The BQ has been validated against a range of other outcomes and PROMs including sick leave, recurrence of pain, disabling pain, persistent pain, the Roland Morris Disability Questionnaire, and the SF-36. A neck version of the Bournemouth Questionnaire has been developed, and both neck and low back versions have undergone translation and testing for cross-cultural validity (Bolton and Humphreys, 2002). In total, the BQ has become available in Dutch (Schmitt *et al.,* 2009; Schmitt *et al.,* 2013), German (Soklic *et* al., 2012; Blum-Fowler *et al.,* 2013), French (Martel *et al.,* 2009), Italian (Geri *et al.,* 2015) and Danish (Hartvigsen *et al.,* 2005).

The BQ is designed for patient completion. Each item is scored on a 0-10 point scale. The scores for each item are summed to produce a total score which can produce a value between a minimum score of 0, and a maximum score of 70. The higher the score reflects the degree of impact on a patient's life. Since its original development, the BQ has been employed in various settings including randomised controlled trials (Gemmell and Miller, 2010; Cramer *et al.,* 2013), feasibility studies (Cheshire *et al.,* 2011), prospective cohort studies (Murphy *et al.,* 2006; Langworthy and Breen, 2007; Murphey *et al*., 2009; Humphreys and Peterson, 2013), predictive studies (Larsen *et al.,* 2005; Peterson *et al.,* 2012), retrospective studies (Dunn *et al.,* 2011), screening studies (Murphy and Hurwitz, 2011), service evaluation (Gurden *et al.,* 2012), and case studies (Rankin, 2006). The systematic review will now address the measurement properties of the BQ.

4.3 Method

4.3.1 The search strategy

A range of search approaches were used for this systematic review. Subscription and free-to-access databases were searched. Hard copies of osteopathic, physiotherapy, and chiropractic journals were searched also. The databases searched are listed in Table 4.3.

Table 4.3 Databases searched for the review

Database	Description	Dates searched
Medical Literature Analysis and Retrieval System Online (MEDLine) and pre-MEDLine	Journals of the United States National Library of Medicine	1966 to present
Allied and Complementary Medicine Database (AMED)	Allied and Complementary Medicine	1985 to present
Web of Science (formerly ISI Web of Knowledge)	Scientific citation indexing service, and access to scientific journals	Inception to present
EMBASE (Excerpta Medica database)	European database of biomedical literature	1947 to present
CINAHL (Cumulative Index to Nursing and Allied Health Literature)	Nursing and allied health research	Inception to present
LILACS (Literatura Latino Americana em Ciências da Saúde)	Latin American and Caribbean Health Sciences Literature	Inception to present
PsycInfo	Abstracts in psychology produced by the American Psychological	Inception to present

	Association	
CENTRAL	Cochrane Central Register of Controlled Trials	Inception to present
Cochrane database of systematic reviews	Systematic reviews in healthcare	Inception to present
clinicaltrials.gov	US Registry of ongoing clinical trials involving human participants	Inception to present
Ovid HealthStar	Published literature on health services, technology, administration, and research	Inception to present
Pascal	European Science, Technology and Medicine database	Inception to present
Physiotherapy Evidence Database (PEDro)	Database of randomised trials, systematic reviews and clinical practice guidelines in physiotherapy.	Inception to present
DIMDI (German Institute of Medical Documentation and Information)	Database-supported information systems	Inception to present
OSTMED.DR	Osteopathic Medicine Digital Repository	Inception to 2003, and 2006 to present
Index to Chiropractic Literature	indexing of the peer-reviewed literature produced by chiropractic publishers	Inception to present
Osteopathic Research Web	Database of Vienna School of Osteopathy and	Inception to present

	UK Osteopathic Educational Institutions	
System for Information on Grey Literature in Europe (SIGLE)	Grey literature database	1980 to present
Osteopathicresearch.net	The Clinical Research Database of the American Association of Colleges of Osteopathic Medicine (AACOM)	Inception to present

4.3.2 Search strings

The search strings used are based on a strategy used by Schellingerhout *et al.*, 2012. This strategy is, in turn, based on a strategy originally devised by Terwee *et al.* (Terwee *et al.*, 2009a). Search strings used in Medline (Ovid) filter used by Terwee *et al.* were adapted by Schellingerhout *et al.* to encompass additional databases including EMBASE and PsycInfo: these were supplied after personal correspondence with the author (Terwee *et al.*, 2009a; Schellingerhout *et al.*, 2012). The search strings are detailed in Appendix 4.1.

Other databases

Key search words used in the additional databases listed include: patient reported outcome measure*, outcome*, back pain, spinal pain, musculo?skeletal, reliability, validity, responsiveness, instrument*, and measurement propert*. The search strategy as shown above makes use of "wildcards" *e.g.* to incorporate singular and plural versions of words when searching. Truncation was also used in the initial scoping search for papers; this process makes use of the symbols "#" or "?" in the middle of a word to either search for several words within a set of letters *e.g.* ne#t to find all citations containing neat, nest or next, or to accommodate variations in spelling between different versions of English *e.g.* color AND colour. Reference lists were examined from the papers identified in the search. Specific author searches were used to include key researchers in the area of outcome research including Bolton J, Breen A, Fairbank J, Roland M, Waddell G, Williams NH, Morris

R, Garratt A, Fitzgerald R, Stanisziewska S, Kravitz R, Pincus T, Ware J, Litcher-Kelly L, Müeller U, and Haywood K, Grotle M, Terwee C, Mokkink L, Ostelo R, Froud R, and de Vet H. Personal databases, and peer databases were searched also to identify additional papers of potential relevance. Clear inclusion and exclusion criteria were identified to inform the review. These are listed in Table 4.4.

Table 4.4 Systematic review inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Studies of validity involving the Bournemouth Questionnaire.	Randomised controlled trials, non-randomised controlled trials, retrospective cohort studies solely investigating the use of medication alone in the management of musculoskeletal symptoms.
Studies of reliability involving the Bournemouth Questionnaire.	Surveys, case studies, editorials, case series, opinion pieces, commentaries, and literature reviews.
Studies of responsiveness involving the Bournemouth Questionnaire.	Trials not involving spinal symptoms.
Studies involving translation, and cross-cultural validation involving the Bournemouth Questionnaire.	Trials involving patients with paralysis.
Studies involving adults (age 18 years and over).	Cadaveric studies.
Studies in English.	Animal studies.

Searches were conducted in 2014, and updated at the end of 2016 to check for new publications.

4.3.3 Analysis, and assessment of articles

Once all searches had been completed, they were exported to an Endnote™ database (Adept Scientific, UK: version 17.3.1.8614). This database was searched using author and titles to identify duplicates which were removed. In situations

where duplicates of the same studies appeared under different titles, the initial version of the paper was retained. Outcome measures have many synonyms and different abbreviations to describe the same measure, and this was accommodated in the search and examination of the number of "hits" (*i.e.* the identification of studies from database and other searches). Studies were reviewed independently by DC and CF. Findings were compared and any areas of disagreement were resolved by discussion in the first instance with the option of referring to the second supervisor (RF) for further opinion. Identified studies were initially assessed in a logical sequence based on the criteria listed in Table 4.5.

Table 4.5 Criteria for rejection of studies

- Were the studies in English?
- Did the studies include musculoskeletal care?
- Did the studies involve one or more patient reported outcome measures?
- Did the study involve patients with paralysis or back pain due to pathology?
- Was the trial a duplicate publication of a trial?

Where insufficient information was available to make an assessment, full text versions of papers were obtained.

Data extraction

Data were extracted to determine:

- Study citation;
- Population in which the study was conducted;
- The sample size of the studies;
- The setting;
- The country in which the research took place;
- The outcomes used.

Quality appraisal

The measurement properties of the papers were assessed using the COSMIN

checklist. This was undertaken separately by both myself and DC, we then conferred to achieve consensus. Within the included studies, concepts including validity, reliability, and responsiveness were assessed. The COSMIN taxonomy includes the following properties which are assessed in the quality appraisal tool:

▪ Reliability	Internal consistency	
	Reliability	Test-retest
	Measurement error	Test-retest
▪ Validity	Content validity	Face validity
	Criterion validity	Concurrent validity
		Predictive validity
	Construct validity	Structural validity
		Hypotheses-testing
		Cross-cultural validity
▪ Responsiveness		

The COSMIN checklist utilises different criteria and statistical methods to assess the measurement properties of different PROMs. These include the items described in Table 4.6.

Table 4.6 – Summary of assessment criteria from COSMIN tool (Mokkink *et al.*, 2010)	
Measurement property	Criteria for assessment
Internal consistency	Presence of effect indicators in the scale. *Design requirements* – percentage of missing items, and how they were handled; - Adequacy of the sample size; - Checking of the unidimensionality of the scale (using factor analysis or application of IRT model); - Calculation of a statistic for internal consistency for each subscale separately; - The presence of any important design or methodological flaws. *Statistical methods* - Calculation of Cronbach's alpha for Classical Test Theory (CTT), continuous scores; - Calculation of Cronbach's alpha or KR-20 for CTT, dichotomous scores; - Calculation of goodness of fit e.g. χ^2, for IRT.
Reliability	*Design requirements* - Percentage of missing items, and how they were handled; - Adequacy of the sample size; - Availability of at least two measurements; - Independent administration; - Statement of time interval used; - Stability of the symptoms of included participants; - Appropriateness of the time interval;

		- Similarity of the test conditions; - The presence of any important design or methodological flaws.
	Statistical methods	- Calculation of an intraclass correlation coefficient (ICC) for continuous scores; - Calculation of kappa score for dichotomous/nominal/ordinal scores; - Calculation of weighted kappa score for ordinal scores; - Description of the weighting scheme for ordinal scores e.g. linear, quadratic.
Measurement error	*Design requirements*	- Percentage of missing items, and how they were handled; - Adequacy of the sample size; - Availability of at least two measurements; - Independent administration; - Statement of time interval used; - Stability of the symptoms of included participants; - Appropriateness of the time interval; - Similarity of the test conditions; - The presence of any important design or methodological flaws.
	Statistical methods	- Calculation of the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC), or Limits of Agreement for CTT.

Content validity	*General requirements* - Assessment of whether all items refer to relevant aspects of the construct to be measured; - Assessment of whether all items are relevant for the study population; - Assessment of whether all items are relevant to the measurement instrument and its intended purposes; - Assessment of whether all constructs comprehensively reflect the construct measured; - The presence of any important design or methodological flaws.
Structural validity	Presence of effect indicators in the scale. *Design requirements* – percentage of missing items, and how they were handled; - Adequacy of the sample size; - Checking of the unidimensionality of the scale (using factor analysis or application of IRT model); - Calculation of a statistic for internal consistency for each subscale separately; - The presence of any important design or methodological flaws. *Statistical methods* - Performance of confirmatory factor analysis for CTT; - Performance of IRT tests for determining the uni-dimensionality of items measured.
Hypotheses testing	*Design requirements* – percentage of missing items, and how they were handled; - Adequacy of the sample size;

		- Formulation of hypotheses *a priori* regarding correlations or mean differences; - Inclusion of the expected direction of correlations of mean differences in the hypotheses; - Inclusions of the expected absolute or relative magnitude of correlations or mean differences in the hypotheses; - Provision of an adequate description of the comparator instrument(s) used for convergent validity; - Provision of an adequate description of the measurement properties of the comparator instrument(s) used for convergent validity; - The presence of any important design or methodological flaws.
	Statistical methods	- Adequacy of the design and statistical methods for the hypotheses to be tested.
Cross-cultural validity	*Design requirements*	– Percentage of missing items, and how they were handled; - Adequacy of the sample size; - Description of the language of the original instrument, and transcribed version; - Adequacy of description of the expertise of the people involved in the translation process; - Independent working of the translators; - Conduct of forward and backward translation of items; - Adequacy of description of how differences in translated versions were resolved; - Review of translation by committee; - Pre-testing of translated version to check interpretation, cultural relevance, and ease of

		comprehension; - Adequate description of the pre-test sample; - Similarity of samples for all characteristics except language and/or cultural background; - The presence of any important design or methodological flaws.
	Statistical methods	- Performance of confirmatory factor analysis for CTT; - Assessment of differential item function (DIF) between the two language groups.
Criterion validity	*Design requirements*	– Percentage of missing items, and how they were handled; - Adequacy of the sample size; - Consideration of the criterion used as a reasonable "gold standard"; - The presence of any important design or methodological flaws.
	Statistical methods	- Calculation of correlations, or the area under the receiver operating curve AUC for - - continuous scores; - Determination of sensitivity and specificity for dichotomous scores.
Responsiveness	*Design requirements*	- Percentage of missing items, and how they were handled; - Adequacy of the sample size; - Availability of at least two measurements within a longitudinal design; - Statement of time interval used; - Adequate description of any events which occurred in the interim period (e.g.

	intervention, or other relevant events); - Recorded change in a proportion of the patients (either deterioration or improvement) *Design requirements for hypotheses testing- Where a "gold standard" was not available for some constructs* - A priori formulation of hypotheses about changes in scores; - Inclusion of direction of correlation or mean differences in hypotheses; - Inclusion of expected absolute or relative magnitude of correlations or mean differences of change scores in the hypotheses; - Adequacy of description of comparator instrument; - Adequacy of description of measurement properties of comparator instrument; - The presence of any important design or methodological flaws. *Design requirements for comparison to an available "gold standard"* - Consideration of the gold standard as a reasonable criterion for change (an provision of evidence in support); - The presence of any important design or methodological flaws. *Statistical methods* - Calculation of correlations between change scores, or the ROC Curve (AUC) for continuous scores; - Determination of sensitivity and specificity for dichotomous scores.

The next section will concentrate on the findings of the search, and the quality assessment of the identified papers.

4.4 Results

4.4.1 The search

The results shown in Table 4.7 include the findings or "hits" achieved from searching each database and research resource.

Table 4.7. Findings from database searches

Updated search May, 2015	Database	Number of hits	Relevant hits
	Index to chiropractic literature	142	28
	ISI Web of Science	144	129
	LILACS	61	1
	www.Clinicaltrials.gov	65	32
	DIMDI	115	2
	PEDRO	156	46
	EMBASE	496	116
	PubMed	4236	892
	PsycInfo	345	21
	Cochrane database	4	2
	SIGLE	23	1
	OstMed (http://www.ostmed-dr.com:8080/vital/access/manager/Index)	59	1
	Osteopathic research web (http://www.osteopathicresearch.org/)	0	0
	Other (reference lists and author searches)	22	53
	Combined hits	6264	1324
	After removal of duplicates	6243	926

Once searches from all databases were combined, they produced a total of 6265 article titles to be screened. Papers were reviewed and either included or excluded systematically. These stages are summarised according to the PRISMA guidelines flowchart (Moher *et al.,* 2009) in Figure 4.1. A copy of the PRISMA checklist is in Appendix 4.2.

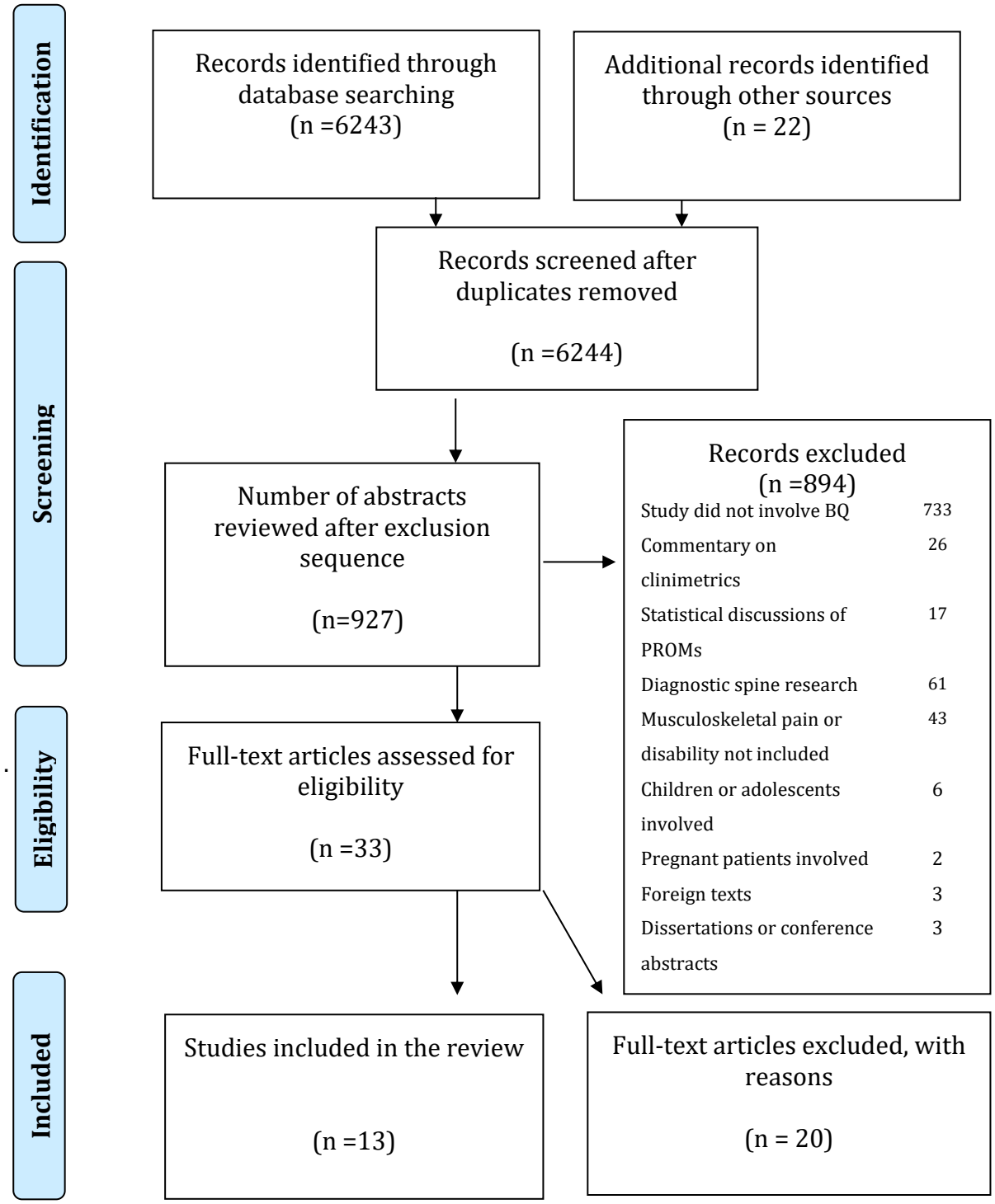


Figure 4.1 Flow chart showing search results

Examination of titles identified some that were immediately irrelevant, and they were excluded from further examination. Some articles were laboratory-based only, and concerned with testing laboratory equipment rather than focussing on feedback from patients. The search strings used were extremely sensitive: even after exclusion of irrelevant studies many remained which required further examination. A total of nine hundred and twenty six abstracts were reviewed independently by CF and DC. From this number, eight hundred and ninety five were excluded for a range of reasons as described in Table 4.8.

Table 4.8. Reasons for rejection of excluded articles

Reasons for rejection	Number of articles
Study did not involve BQ	733
Commentary on clinimetrics	26
Statistical discussions of PROMs	17
Diagnostic spine research	61
Musculoskeletal pain or disability not included	43
Children or adolescents involved	6
Pregnant patients involved	2
Foreign texts	3
Dissertations or conference abstracts	3

Full text versions were retrieved for 32 studies, with 13 studies being selected finally for inclusion in the review. The characteristics of the included studies were summarised based on author(s), study population, sample size, setting, country, and outcomes included in the examined studies. This information is summarised in Table 4.9.

Table 4.9 General characteristics of included studies					
Author(s)	Population	Sample size	Setting	Country	Outcomes measures used in study
Bolton and Breen, 1999	New patients or former patients with a new episode of back pain (with or without leg pain)Mean age 52.1 (SD=15.5)	n=90	Single chiropractic practice	England	
	New patients or former patients with a new episode of back pain (with or without leg pain). Mean age 49.5 (SD=16.9)	n=82	Chiropractic college teaching clinic	England	
	New patients or former patients with a new episode of back pain (with or without leg pain). Mean age 45.7 (SD=12.5)	n=55	Field chiropractic clinics	England	Chronic Pain Grade Questionnaire (CPG);Revised Oswestry Disability Questionnaire (RODQ);Distress and Risk Assessment Method (DRAM);Pain Locus of Control (PLC) Scale; Fear-Avoidance Beliefs Questionnaire (FABQ)
Bolton and Humphreys, 2002	Patients with non-specific neck pain	n=102	Chiropractic college teaching clinic and 8 field clinics	England	BQ neck version, Neck Disability Index (NDI), Copenhagen Neck Functional Disability Scale (NFDS), and the Short Form-36 (SF36).
Perillo and Bulbulian, 2003	Consecutive new patients with low back pain from a chiropractic teaching college clinic. Mean age 38.6.	n=70 (37M; 33F)	Chiropractic teaching college clinic	USA	Bournemouth Questionnaire (BQ), Numerical Rating Scale (NRS), Oswestry Disability Index (ODI), Global Rating of Change and Importance of Change.
Hartvigsen *et al.,* 2005	Random sample of low back pain patients for translation of BQ into Danish.	n=30	Two chiropractic clinics.	Denmark	

	New patients with low back pain to assess validity.	n=118	Seven chiropractic clinics.		Pre-treatment version of the BQ (Danish version), SF-36, and RMDQ at initial consultations. At four weeks, patients received the post-treatment BQ (Danish), the SF-36, and the RMDQ.
	Patients with chronic low back pain attending an introductory hospital meeting.	n=28	Outpatient back-pain clinic.		BQ (Danish version) at pre-meeting, and again 2 hours later.
Larsen and LeBoeuef-Yde, 2005	Patients with persistent low back pain (lasting longer than 2 weeks, or experiencing 30 days of low back pain during the previous year	n=875	Chiropractic clinics	Norway	BQ (Norwegian version), revised Oswestry Disability Questionnaire, additional questions on the number of days with low back pain, and the number of days off work during the past year.
Gay *et al.,* 2007	Patients with neck pain. Mean age 49.6 years (SD=14.6).	n=23	Outpatient physical therapy clinic	USA	Neck Disability Index (NDI), neck Bournemouth Questionnaire (NBQ), and a pain Visual Analogue Scale (VAS).
Martel *et al.,*2009	Patients with chronic cervical pain (minimum duration 12 weeks). Mean age 41.1 years (SD=10.1)	n=68 (46F;21M)(sic)	University	Canada	Neck Disability Index (NDI)Visual Analogue Scale (VAS)FABQ-1Bournemouth Neck Questionnaire (French version) (BQc-f)
Schmitt *et al,,* 2009	Patients with subacute, and chronic Whiplash Associated Disorder (WAD). Mean age for F=41.8 (11.7), and mean age for M=44.5(11.1).	Cross-cultural adaptation into Dutch n=92 (69F; 23M)	Patients recruited via GPs, emergency departments, and WAD support website.	The Netherlands	Dutch version of the neck Bournemouth Questionnaire (NBQ-NL), and the Dutch language versions of the Neck Disability Index (NDI), the Hospital Anxiety and Depressions Scale (HADS), the General Perceived Self-efficacy Scale (GPSES), the SF-36, and the pain Numerical Rating Scale (NRS-pain).
	Patients with subacute	Test-retest reliability, n=34			NBQ-NL

	Whiplash Associated Disorder (WAD).				
Newell and Bolton, 2010	Patients undergoing chiropractic treatment for acute and subacute/chronic low back pain	n=437(Subacute/chronic: n=251)(Acute: N=186)	Chiropractic clinic	England	Bournemouth Questionnaire Patient Global Impression of Change (PGIC)
Blum-Fowler *et al.,* 2013	Patients with low back pain	Face validity: n=30	Not disclosed	Switzerland	Bournemouth Questionnaire for low back pain Bournemouth Neck Questionnaire (German version)Oswestry Disability Index (German version) (ODI)Short Form-36 (SF-36)
	Students with low back pain	Test-retest reliability: n=30	Eidgenösische Technische Hochshcule, Zurich		
	Patients with low back pain	Internal consistency: n=108	Five chiropractic clinics		
Schmitt *et al.,* 2013	Physiotherapists attending a CPD workshop. Mean age 35 years (range = 24-63)	n=22 (14F; 8M)	Not disclosed	The Netherlands	Neck Bournemouth Questionnaire – Dutch version (NBQ-NL)
Soklic *et al.,* 2013	Patients with neck pain	Translation and cross-cultural adaptation n=30	Chiropractic clinic	Switzerland	Pre-final German version of the Bournemouth neck questionnaire (BQN)
	Medical students and chiropractors	Test-retest reliability n=31	Lecture for medical students, and professional meeting for chiropractors	Switzerland	German version of the Bournemouth neck questionnaire (BQN)
Geri *et al.,* 2015	Participants with neck pain of mean duration	n=180 (80F;28M)	Outpatient physiotherapy	Italy	Neck Bournemouth Questionnaire (NBQ)Neck Pain and Disability Scale

	12.3 months (SD7.5)		service		(NPDS)European Quality of Life 5 Dimensional Scale (EQ5D)Numerical Rating Scale for Pain Intensity (NRS-Pain)Global Perceived Effect (GPE)

4.4.2 Reporting methods

In summary, a range of measurement properties were examined in the selected studies including two or more within one study. These included:

- Validity;
- Test-retest reliability;
- Responsiveness;
- Translation and cross-cultural validity.

Although guidelines exist to support clinimetric evaluation of patient reported outcome measures, a series of different methodological approaches have been used to assess the different versions of the Bournemouth Questionnaire. A summary of the statistical evaluations is included in Figure 4.2, and a more detailed summary of information from the studies is included in Table 4.10. The terms described in Table 4.10 are recorded as they appear in the original papers.

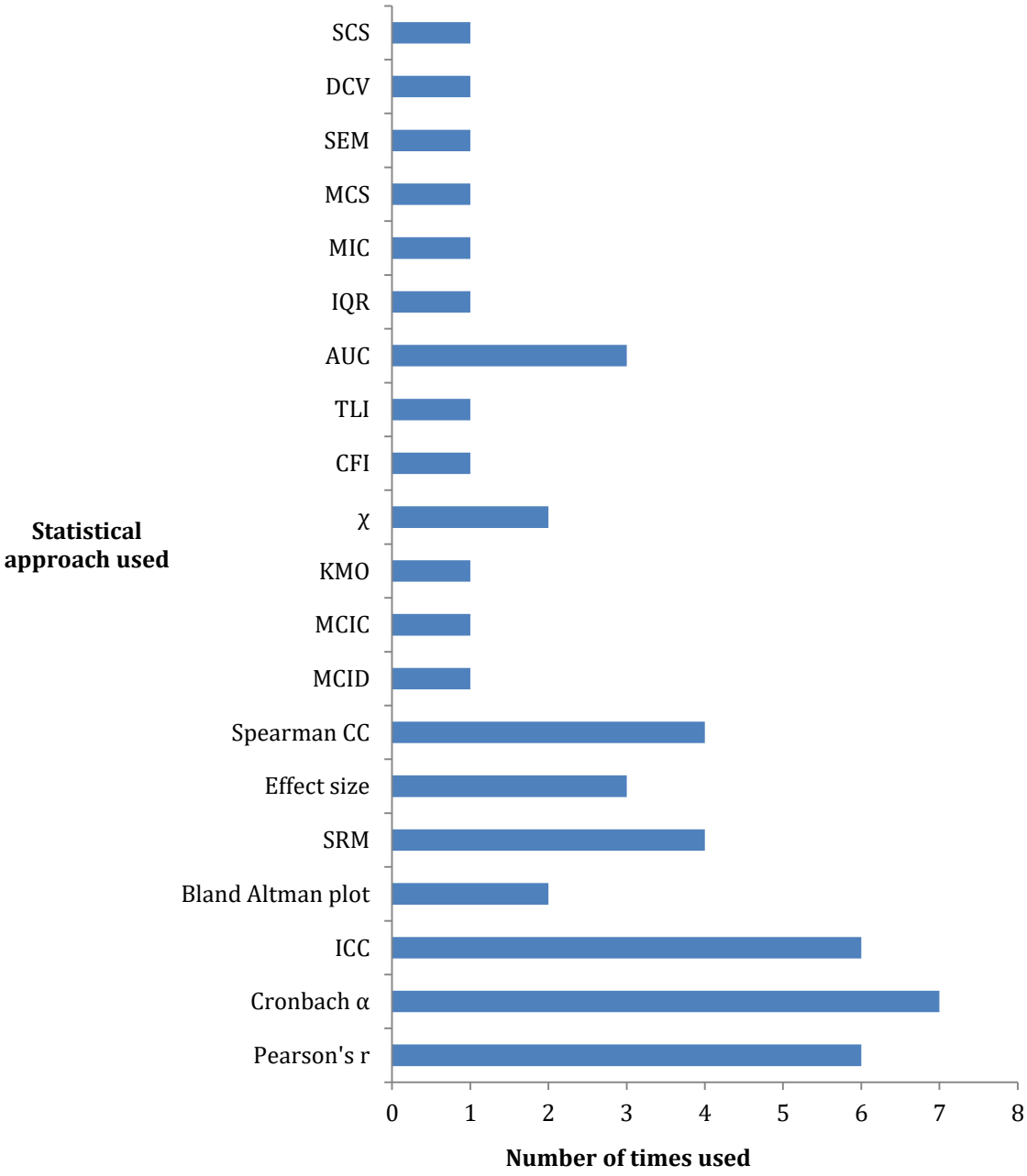


Figure 4.2. Statistical approaches used in BQ clinimetric evaluations

Abbreviations used in Figure 4.2 include:

ICC:	Intra-class Correlation Coefficient	CFI:	Comparative Fit Index
SRM:	Standardised Response Mean	TLI:	Tucker-Lewis Index
KMO:	Kaiser-Meyer-Olnik coefficient	AUC:	Area under the ROC Curve
SCC:	Spearmann Correlation Coefficient	IQR:	Inter Quartile Range
MCID:	minimal clinically important difference	MCS:	Minimum Change Score
MCIC:	minimal clinically important change	SEM:	Standard Error of Measurement
MIC:	minimum important change	DCV:	Discriminant Content Validity
X^2	Chi Square test	SCS:	Standardised Change Score

Table 4.10 Summary of data from included studies													
	Bolton and Breen, 1999	Bolton and Humphreys, 2002	Perillo and Bulbulian, 2003	Larsen and Le Boef, 2005	Hartvigsen *et al.,* 2005	Gay *et al.,* 2007	Martel *et al.,* 2009	Schmitt *et al.,* 2009	Newell and Bolton, 2010	Blum-Fowler *et al.*, 2013	Schmitt *et al.,* 2013	Soklic *et al.,* 2013	Geri *et al.,* 2015
Body area	Lumbar spine	Cervical spine	Lumbar spine pain with or without cervical pain	Lumbar spine	Lumbar spine	Cervical spine	Cervical spine	Cervical spine (WAD)	Lumbar spine	Lumbar spine	Cervical spine	Cervical spine	Cervical spine
Duration of symptoms	New patients	New patients or new episodes of neck pain	> 3 months	> 2 weeks	> 12 weeks	> 12 weeks	> 12 weeks	. 3weeks and , 6 months	<6 weeks; 7-12 weeks; > 13 weeks.	Not stated	N/A	Not stated	> 12 weeks
Measure-ment property investigated	Responsive-ness; Test-retest reliability; validity; Internal consistency.	Test-retest reliability, validity, and responsive-ness	Responsive-ness	Predictive validity	Cross-cultural adaptation; internal consistency; Test-retest reliabilit	Responsiveness	Cross-cultural adaptation; Test-retest reliability; Responsiveness.	Cross-cultural adapta-tion; Test-retest reliability.	Responsiveness	Cross-cultural adaptation Test-retest reliability;	Content validity	Test-retest reliability (TRT); Internal consistency (IC);Construct validity (CV);Responsiveness (R); cross-	Cross-cultural adaptation; internal consistency; Test-retest reliability; Responsiveness

					y; Responsiveness							cultural adaptation.	
Time interval between studies (where relevant)	Same day (TRT);4 weeks (responsiveness)	Pre-treatment, later in day after treatment, and 4-6 weeks post-treatment	Baseline, 15 days, 30 days, 45 days.	Baseline, 4th visit, 3 months, and 12 months.	2 hours (TRT);4 weeks later (R).	4 weeks	24 hours	3 weeks	4 weeks	TRT: 2 hours; Validity: 4 weeks.	N/A	TRT: 2 hours; R: 4 weeks	
Statistical tests used	Effect size; ICC; Pearson r; Cronbach α.	Cronbach's α, ICC, Pearson's r, Mean change score, and effect size.	Standardised change scores; Relative efficiency Minimum clinically important difference (MCID); ROC plots.	ROC values.	ICC; Cronbach α;SRM	Effect size; Cronbach α;SRM.	Effect size; SRM; ICC; Pearson r; MIC.	ICC;SEM; Cronbach α.	SRM:MCIC	ICC; Cronbach α;SRM.	ICC	SRM; ICC; Cronbach α.	Cronbach α;MCID;X^2;R;AUC.

Values obtained	Effect size: 1.29; ICC: 0.95 (total score); Pearson r: 0.56-0.71; Cronbach : 0.9.	Cronbach's α: 0.9, ICC: 0.65, Pearson's r:0.42-0.82 , Mean change score: 22.8, and effect size: 1.67.	SCS: 0.78MCID: 5	ROC values for predictive validity of sum scores of the two questionnaires were for revised Oswestry Questionnaire were 0.56-0.63, and 0.56-0.62 for the BQ	ICC:0.96; Cronbach α: 0.89; SRM: Pain:1.82 Phys funct:1.32 Social funct:0.90 Anxiety: 0.95 Depress:0.76 Pain control: 0.91	Effect size: 1.28; Cronbach α: 0.85 (pre-treatment) and 0.89 (post-treatment); SRM: 1.17.	Effect size: 0.56;SRM: 0.61;ICC: 0.97 (0.95-0.98);Pearson r (TRT): 0.97 (0.95-0.98);Pearson r:0.61(0.44-0.74) pre-construct validity;0.67 (0.51-0.78) post-construct validity; MIC: 4.4 points.	ICC; 0.83-0.92Cronbach α: 0.87 (sum scores)SEM: 3.67(real change above measurement error would be 10.2 points	SRM: Improved acute patients: 1.9 (1.7-2.0) Non-improved acute patients: 1.2 (0.9-1.5);Improved subacute/chronic patients: 1.7 (1.5-1.8);Non-improved subacute/chronic patients: 0.5	ICC:>0.91Cronbach α: 0.86-0.94 (pre-treatment to post-treatment)	ICC: Q1: 0.953Q20.944Q30.955Q40.963Q50.975Q60.876Q70.952	ICC: (sum score: 0.99)Q1: 0.98Q2: 0.91Q3: 0.92Q4: 0.93Q5: 0.97Q6: 0.96Q7: 0.96Cronbach α: 0.79	Cronbach α: 0.89; MCID: 5.5 points (sensitivity=0.75:specificity=0.60);AUC: 0.72; X^2:R: 0.69.

									(0.3-0.7)MCIC: acute patients: 26 points; Subacute patients: 18 points.				

ICC:	Intra-class Correlation Coefficient	CFI:	Comparative Fit Index
SRM:	Standardised Response Mean	TLI:	Tucker-Lewis Index
KMO:	Kaiser-Meyer-Olnik coefficient	AUC:	Area under the ROC Curve
SCC:	Spearmann Correlation Coefficient	IQR:	Inter Quartile Range
MCID:	minimal clinically important difference	MCS:	Minimum Change Score
MCIC:	minimal clinically important change	SEM:	Standard Error of Measurement
MIC:	minimum important change	DCV:	Discriminant Content Validity
X^2	Chi Square test	SCS:	Standardised Change Score
ROC	Receiver Operator Characteristic		

Once the characteristics of the included studies had been examined, they were assessed individually using the COSMIN quality appraisal tool. This was undertaken individually by CF and DC, and the findings were discussed to reach a consensus for each relevant measurement property included in the studies. The consensus findings are summarised in Table 4.11.

Table 4.11 Methodological quality of each study per measurement property

Author and date	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Responsiveness
Bolton and Breen, 1999	Fair	Good	N/A	Good	Fair	N/A	N/A	Excellent	Good
Bolton and Humphreys, 2002	Good	Good	N/A	Excellent	Good	Good	N/A	Good	Good
Perillo and Bulbulian, 2003	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Good
Larsen and LeBoeuf, 2005	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Fair
Hartvigsen *et al.,* 2005	Fair	Fair	N/A	N/A	N/A	N/A	Good	N/A	Fair
Gay *et al.,* 2007	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Poor
Martel *et*	Fair	Fair	N/A	N/A	Fair	N/A	Fair	N/A	Good

al., 2009									
Schmitt *et al.,* 2009	Good	Good	N/A	N/A	N/A	N/A	Good	N/A	N/A
Newell and Bolton, 2010	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Good
Soklic *et al.,* 2012	Fair	Good	N/A	Good	Good	N/A	Good	N/A	Fair
Blum-Fowler *et al.,* 2013	Good	Fair	N/A	Good	Good	N/A	Excellent	N/A	N/A
Schmitt *et al.,* 2013	N/A	N/A	N/A	Excellent	N/A	N/A	N/A	N/A	N/A
Geri *et al*., 2015	Excellent	Good	N/A	Excellent	Good	Excellent	Excellent	N/A	Good

This systematic review examined 13 studies in total published between 1999 and early 2015: they included evaluation of different measurement properties, and translations into five European languages including Danish (Hartvigsen *et al*., 2005), Italian (Geri *et al*., 2015), Dutch (Schmitt *et al*., 2009; Schmitt *et al.,* 2013), German (Soklic *et al*., 2012; Blum-Fowler *et al*., 2013), and French (Martel *et al*., 2009). The studies included patients with neck pain (n=7) and low back pain (n=56. The populations involved included medical students, chiropractors, physiotherapists, and patients attending clinics with either neck or low back pain.

Among the studies involving the cervical spine (neck), five involved cross-cultural translations (Martel *et al*., 2009; Schmitt *et al*., 2009; Soklic *et al.,* 2012; Schmitt *et al*., 2013; Geri *et al*., 2015), one evaluated measurement properties (Bolton and Humphreys, 2002), and one undertook a comparison of the BQ with the Neck Disability Index or NDI (Gay *et al.,* 2007). Patients involved in the studies had experienced symptoms from 6 weeks to 6 months: although the time interval between testing varied depending on the measurement property being evaluated. In the studies examining cross-cultural translation, the time interval between administration of the BQ ranged from two hours to three weeks. Among the studies examining low back symptoms, two examined cross-cultural translations (Hartvigsen *et al*., 2005; Blum-Fowler *et al*., 2013), and four examined other measurement properties (Bolton and Breen, 1999; Bolton and Humphreys, 2002; Perillo and Bulbulian, 2003; Larsen and Leboeuf-Yde, 2005; Newell and Bolton, 2010).

Translation and cross-cultural adaptations can be performed based on guidance from Beaton *et al.*, 2000. The studies involving the Bournemouth Questionnaire (BQ) involved a minimum of three stages using independent translators, review of the translation by the original translators with oversight by an expert committee, and finally back translation into English by translators with English as their first language but fluency into the translation language (Soklic *et al*., 2012). In other translation studies, six stages were present including the additional stages of expert validity, face validity in patients, and final audit of the process by an expert

group (Hartvigsen *et al*., 2005; Schmitt *et al*., 2009; Blum-Fowler *et al*., 2013; Geri *et al*., 2015).

I will now summarise the main findings of this review in relation to my original objectives before I contrast the findings of this study with reviews of other PROMs. I will conclude by discussing the strengths and limitations of this review, and give a brief critique of my experience of using the COSMIN evaluation tool.

Reliability

This was evaluated in eight of the studies. Five were evaluated as good, and three as fair. Values for ICC ranging from 0.65 to 0.95 were reported. Internal consistency was evaluated in eight studies: four were evaluated as fair, three as good, and one as excellent. Reported values for Cronbach α ranged from 0.85 to 0.94.

Translation and cross-cultural validity

Translation occurred in five studies and involved a selection of European languages including French, Dutch, Italian, and French. The neck version of the BQ was more commonly translated. The translation process followed a consistent pattern across studies. One of the studies was evaluated as fair; two as good; and two as excellent.

Responsiveness

Nine studies provided an evaluation of responsiveness, although the manner in which this was reported varied across studies. Six studies were evaluated as good; two as fair; and one as poor. Reporting statistics involved the use of effect size (0.56-1.67), and the area under the curve (AUC) which was reported as 0.72 in a single study. A detailed description of the findings and statistics for all of the studies reviewed is shown in Table 4.10 (statistics) and Table 4.11 (evaluation using the COSMIN checklist).

One study examined whether the BQ could be used to monitor and predict treatment outcome in patients presenting with symptoms of persistent low back

pain Larsen and Lebouef-Yde, 2005). Patients (n=875 were included if they reported symptoms for 2 weeks at baseline or 30 days within the preceding year. The BQ amd ODI were completed by patients. The authors concluded that the BQ was not a useful instrument to identify baseline status, monitor treatment progress, or predict progress 1 year post treatment conclusion. However, the autors did note that questions relating to pain cnytrol, activities of dailiy living, and fear avoidance in relation to work activities were useful to predict outcome. This study stands alone in providing a negative view of the value of the BQ in measuring outcome for monitoring and benchmarking in patients with low back pain.

Statistical evaluation of common measurement properties for the BQ varied across studies, and was consistent with changing expectations from publication. In studies evaluating responsiveness Effect Size (ES), Standardised Change Scores (SCS), Standardised Response Mean (SRM), Minimally Important Change (MIC), Minimally Important Difference (MID), and Receiver Operating Characteristic (ROC) were reported. Validity was reported in studies using either Pearson's or Spearman's correlation coefficient, and reliability was reported using Cronbach's α, or Intraclass Correlation Coefficient (ICC). The statistical tests with their respective values are listed in Table 4.10.

4.5 Discussion

4.5.1 Main findings

This is the first systematic review of the measurement properties of the BQ. Despite using a highly sensitive search string, no other systematic reviews of the BQ have been identified. In summary, this study found:

- The measurement properties of all included studies ranged from excellent to poor when evaluated using the COSMIN checklist;
- Cronbach α is the commonly reported statistic for internal consistency;
- Effect size and standardised response mean are the most commonly reported statistics for responsiveness;
- Intraclass Correlation Coefficient is the most commonly reported statistic to evaluate reliability;

- A consistent and thorough approach has been adopted in all cross-cultural adaptations of the BQ.

4.5.2 Comparison to existing research

Although this is the first systematic review of the measurement properties of the BQ alone, it is illustrative to compare my findings to other commonly used PROMs *e.g.* the Oswestry Disability Index (ODI) and the Roland Morris Disability Questionnaire (RMDQ) as cited by Müller *et al*. as the most commonly used PROMs in the management of patients reporting low back pain (Müller *et al*., 2004).

Several reviews of the measurement properties of the ODI and RMDQ exist (Somerville *et al.,* 2008; Carreon *et al*., 2008; Artus *et al*., 2011; Lin *et al*., 2011; Spanjer *et al*., 2011; Goertz *et al*., 2012; Murphy and Lopez, 2013; Sodha *et al*., 2012; Newman *et al*., 2013; Kamper *et al*., 2015; Oner *et al*., 2016; Yao *et al*., 2016), but only one has used the COSMIN tool for evaluation of the 24-item RMDQ and ODI version 2.1a (Chiarotto *et al*., 2016). The review by Chiarotto *et al*. included nine articles in their review concluding that the ODI displayed smaller measurement error and better test: re-test reliability than the RMDQ. The RMDQ, however, was regarded to show better construct validity. Neither PROM was regarded as having conclusive evidence of good responsiveness or internal consistency. In comparison, this review identified a more consistently sound performance with measures scoring fair or above in relevant measurement properties.

When considering PROMs used in neck pain, the review by Schellingerhout *et al*., 2012 employed the COSMIN tool to evaluate the BQ. Schellingerhout *et al.* reviewed the clinimetric properties of the BQ (neck questionnaire), the Neck Disability Index (NDI), the Neck Pain and Disability Scale (NPDS), the Northwick Park Neck Pain Questionnaire (NPNPQ), the Copenhagen Neck Functional Disability Scale (CNFDS). The Whiplash Disability Questionnaire (WDQ), the Core Neck Questionnaire (CNQ), and the Core Whiplash Outcome Measure (CWOM). At the time of the review in 2012, Schellingerhout *et al.* concluded there were no methodologically sound studies available for the BQ which measured reliability,

internal consistency, measurement error or structural validity. Their inclusion of only three studies up to the search cut-off date of 2010 is surprising. The reviewers, however, did identify positive evidence for hypothesis testing and responsiveness. In this particular review, the research team concluded that the Neck Disability Questionnaire showed the most positive results for internal consistency, content validity, structural validity, hypothesis testing and responsiveness, but negative results for reliability. In contrast, this systematic review has included more studies and recently published studies which challenge these findings particularly for internal consistency and reliability

4.5.3 Strengths and Limitations of this systematic review

This review is the first investigating the measurement properties of the neck and back BQ. While it can be compared to systematic reviews of other low back and neck PROMs, there are no reviews of similar studies involving the BQ alone. The review aimed to provide data on key measurement properties including reliability, responsiveness, and validity.

Typically in any research study or review there will be areas for improvement if the study was repeated. Despite using extensive searches and sensitive search strings, it is possible that some studies have not been identified for consideration in the review. The exclusion of foreign language papers from the review may have deprived the review of additional cultural input in cross-translated studies which have not appeared in English. Exclusion of foreign language papers can be said to introduce a form of bias known as "The Tower of Babel Bias" (Grégoire *et al*., 1995). However, the majority of studies are published in English, or the option of English translations is available from some journal web sites. The extensive time and cost involved in seeking translations of foreign language papers would not have been balanced by the potential impact on this review. It is unlikely, in my assessment, that this particular exclusion criterion has had an effect on the findings of this review. Although there were two reviewers for the studies, I had intended that the searches would be conducted independently also to ensure maximum possibility of identifying all relevant studies.

Moreover, the independent assessment of the included papers was extremely valuable when using the COSMIN tool. There are considerable areas of the tool which require interpretation and subjective evaluation, and the opportunity to be able to discuss the findings from using the tool helped to mitigate the potentially slanted evaluation that may have occurred with a single reviewer.

4.5.4 Implications

The review of the BQ identified that it demonstrates sound measurement properties for inclusion in an electronic data collection facility. The developmental setting of the BQ within private manual therapy practice suggests it is suitable for the clinical setting of interest in this thesis. Notwithstanding the fact that head-to-head studies of paper and electronic assessment of the BQ do not exist, there is the opportunity to assess its measurement properties when used in an electronic format through later studies in this thesis. Bishop *et al*. compared the performance of the RMDQ in electronic and paper versions, but this is one of the few PROMs that has undergone such testing (Bishop *et al*., 2010). The evaluation of so-called "e-PROMs" is currently neglected in the literature (Stone *et al*., 2003; Froud, 2008).

I reported in Table 4.11, that the assessment of the measurement properties of the BQ provided evaluations ranging from poor to excellent. One of the requirements of the COSMIN tool is for sample sizes to be between 50-99, or ≥100 per analysis for the assignment of good or excellent respectively to a measurement property. While the size of studies using manual therapy techniques and the BQ are growing (Hoehler *et al*., 1981; Andersson *et al*., 1999; Licciardone *et al*., 2016), many of the sample sizes in the included studies were below the sample sizes required by COSMIN but were nonetheless robust in other methodological considerations. There is a slight dissonance between the sample sizes indicated in the COSMIN tool to achieve an "excellent" or "good" evaluation, and the recommendations given in other COSMIN resources (Terwee *et al.,* 2012). Larger study sizes in future research using the BQ will contribute to a more accurate evaluation of its measurement properties.

4.5.5 Critique of the COSMIN tool

Evaluating Patient Reported Outcome Measures among the vast array of PROMs there are those that are disease specific and condition specific. In 2002, Garratt *et al*. reported that 1275 PROMs existed having a multitude of applications in terms of their ability to be decision-making, monitoring, or economic evaluative tools (Garratt *et al.,* 2002). Kirshner and Guyatt distinguished three kinds of health status measures including discriminative, predictive, and evaluative measures (Kirshner and Guyatt, 1985). When starting to use PROMs, some of the key issues are where to locate the tools, identifying what are the most suitable tools for a particular population or setting, and have the tools been developed to a sufficiently high standard to deliver robust data. The choice of tool will focus around the concept to be measured, the proposed use of the instrument, and the associated costs, for example (Mokkink *et al.,* 2006).

The task of identifying standards for the development and evaluation of PROMs has been addressed by a series of groups during the past 40 years. While standards for the development of PROMs have been recognised based on the work of McDowell, and Chassany *et al*., standards for evaluation came slightly later (McDowell, 1987; Chassany *et al.,* 2002). The work in developing criteria for evaluation was considered initially by the Scientific Advisory Committee (SAC) of the Medical Outcomes Trust (MOT). The MOT is a non-profit public service organisation established in 1994 to act as a depository and distributor of high quality standardised health assessment instruments (Perrin, 1995; SAC, 1995). The SAC evaluates every instrument in the MOT library according to eight key attributes including the conceptual and measurement model; reliability; validity; responsiveness; interpretability; respondent and administrative burden; alternative forms; and cultural and language adaptations (Lohr *et al.,* 1996). Other initiatives followed to create review criteria including the GraQol Index, 12 standards listed by Bombardier and Tugwell, and standards specifically for disability outcomes (Bombardier and Tugwell, 1987; Andresen and Meyers, 2000; Badia and Baro, 2001).

Valderas *et al*. further developed the criteria of the MOT to create a new tool entitled Evaluating the Measurement of Patient Reported Outcomes or EMPRO (Valderas *et al.,* 2008b). The EMPRO tool was developed using a panel of four experts nominated based on their experience with the development, assessment, and use of outcome measures. The content of the tool was developed based on transforming each of the MOT criteria into individual items, and was formatted according to the Appraisal of Guidelines Research and Evaluation or AGREE instrument (AGREE Collaboration, 2003). A new tool consisting of 39 items was developed and evaluated for feasibility, reliability, validity, and score distributions by 20 reviewers examining Spanish translations of five PROMs *i.e.* SF-36, Nottingham Health Profile, COOP-WONCA charts, EuroQoL-5D, and EORTC-QLQ-C30 (Ware and Sherbourne, 1992; Hunt and McEwan, 1980; Nelson *et al.,* 1987; EuroQol Group, 1990; Aaronsen *et al.,* 1993). It showed good preliminary reliability and validity.

While Valderas *et al*. were developing on the EMPRO tool, work was ongoing in the Netherlands under the aegis of the COSMIN group (Valderas *et al.,* 2008b). They undertook a four-round international Delphi consensus study to develop a checklist to evaluate explicit criteria relating to the measurement properties of PROMs. The criteria were developed from the literature and systematic reviews of measurement properties, and presented to an international panel (n=57) composed of psychologists, clinimetricians, biostatisticians, medical professionals, and members of important organisations involved in instrument evaluation (Mokkink *et al.,* 2010a). The final COSMIN checklist contains ten separate criteria which are answered as part of a four-stage process. There are considerably more questions within each criteria varying depending on what is being assessed. The COSMIN tool has been assessed further to determine the inter-rater agreement and reliability of each item score on the COSMIN checklist (Mokkink *et al.,* 2010b). Good reliability was found but inter-rater agreement was low. The rationale for this was the need for subjective judgements when using the tool, and reviewers being unaccustomed to the different standards, terminology and definitions used by COSMIN. This resulted in further development of the manual which accompanies the tool, and the recommendation for training prior to using the tool.

Notwithstanding the issues identified with the COSMIN tool, it has undergone more extensive development and testing than the EMPRO tool. The COSMIN tool is also used more widely as shown from a Medline search for articles containing both tools (Figure 4.3).

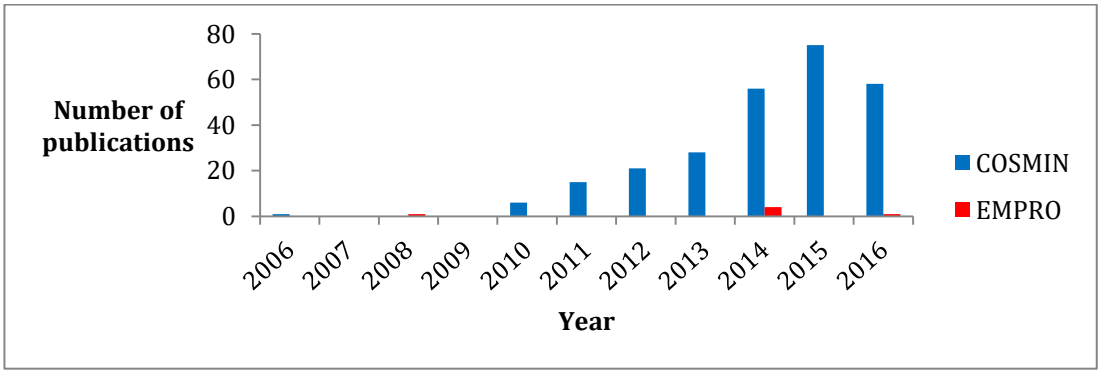


Figure 4.3 Publications using either the COSMIN or EMPRO tools

The number of systematic reviews of measurement properties continues to increase as PROMs' use continues to grow. This is demonstrated by the growth in publications in this area of the literature (Figure 4.4).

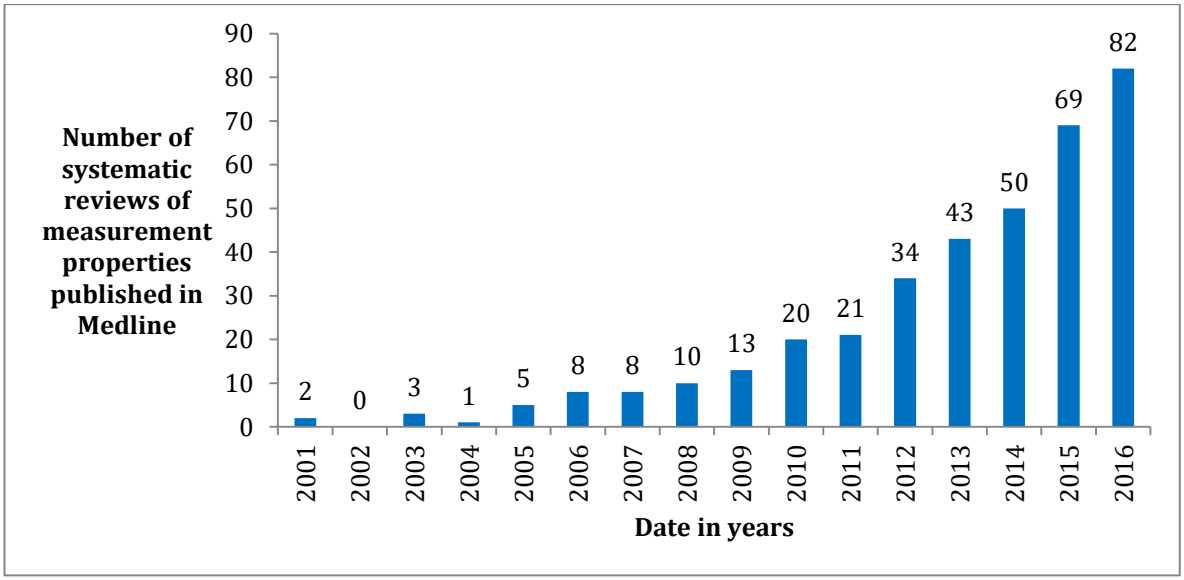


Figure 4.4 Systematic reviews of measurement properties published in Medline 2001 to 2016 inclusive.

When selecting a measurement tool for my systematic review, the COSMIN tool seemed the natural choice. In practice, the tool was far from easy to use. Despite undergoing training, (as recommended by Mokkink *et al.,* 2010b), the tool and manual are extremely complicated and there is a high degree of subjectivity involved. Users of the tool are asked to apply their judgement when evaluating a PROM but this does present issues when trying to attain a standardised approach. Another issue which can be applicable when evaluating any form of publication is the disconnect that can be experienced in the time since publication and the standards required for publication from modern PROMs. Training took place using PROMs published within the past five years, but many of the PROMs used frequently for patients with musculoskeletal symptoms were published in the 1980s and 1990s. Publication requirements were different at that time, but evaluation of a PROM using a tool created more than 25years later can present an overly-critical view of a valid and reliable PROM.

Using the COSMIN tool requires also a "rounding down" approach where the lowest criteria among those selected would be used for an overall evaluation of each of the criteria. This presented challenges where one aspect of a section might have been of a lower standard than desirable in contrast to other aspects which had been evaluated as "good" or "excellent". The COSMIN group have announced that the COSMIN checklist is currently being revised based on user experience (http://www.cosmin.nl/other-studies.html COSMIN, 2015), and this will include particular emphasis on content validity with three new boxes being added to the tool (COSMIN, 2016 http://www.cosmin.nl/28/new-cosmin-guidelines-for-content-validity). The openness of the COSMIN group to re-evaluation of the tool is to be applauded.

4.6 Conclusions

The quality of the research reviewed addressing different clinimetric properties varied according to the properties being investigated. There were differences in the sample sizes, populations, settings, and time durations involved. Statistical analysis was largely consistent depending upon the clinimetric property being investigated, and the rationale for the statistical tests selected were well reasoned.

The BQ is being used increasingly by manual therapy professionals, and there is a steady increase in available publications. The use of the COSMIN tool to evaluate the clinimetric properties of the BQ was not without its challenges. The capacity for subjective assessment of measurement properties undermines the conclusions which can be drawn from its use in my view.

While the measurement properties of the BQ are important when considering its use for individual and collective professional groups, there is a lack of qualitative evaluation by patients concerning its use. Ultimately with PROMs, the patient's view of a questionnaire is important in its completion, and it would be helpful to see patients' evaluations of its use in future publications. A common area which has been lacking in the studies in this review and in many others is the completion time for the questionnaire. This would be a useful item of information for future studies especially as PROM completion becomes increasingly embedded in practice and patients wish to have this information.

Overall, this review has identified some useful findings. It has confirmed the value of using the BQ as one of the PROMs in an electronic data capture system. The development and pilot of an electronic data capture system will be described in Chapter 5.

5

Development and pilot-testing of content for a PROM mobile and web application

5.1 Introduction

Although many Patient Reported Outcome Measures (PROMs) have been developed over the past 30 years, their introduction into day-to-day practice, and their use with technology to support such data collection has been relatively recent (Nelson *et al.*, 2015). Paperless data collection is being advocated increasingly to benefit from technological developments, and increase wider access among the population (Hunt, 2013). This chapter will focus on the development of the content for a PROM application (app) for osteopaths for baseline and follow-up data collection, its pilot testing, the findings of that testing, and the analysis of qualitative and quantitative data collected. While the content of the app is important, there are key stages in the development of the software underpinning the app which must be closely observed to produce a system which is robust.

5.1.1 App development and testing

The first mobile phone with Internet connectivity was launched in Finland in 1996 by Nokia (the Nokia 9000 communicator). Phone network service providers progressively developed systems and services conveniently accessible on mobile phones in combination with reducing costs of mobile phone prices. Small screens and key pad operations were made more efficient by the development of the Wireless Application Protocol (WAP), a specific document and networking model for mobile phones (www.inspiredbloggers.blogspot.co.uk/2004/12/brief-history-of-wap_110252445307049372.html). By 2008 more mobile devices were capable of accessing the Internet, and Internet access by this medium had overtaken personal computers (Hillebrand and Friedhelm, 2010).

Apps

A mobile app (short for application software) is a computer program designed to work on either a smartphone, tablet computer, or other mobile device. Apps are usually available for free or to purchase through application distribution platforms (Grifin and Baston, 2001). Increasing availability of developer tools has meant the expansion of mobile apps from information retrieval systems to games, factory automation, Global Positioning Systems (GPS), order-tracking, and most recently mobile medical apps. Increased popularity and use of apps is producing income generation of US$26 billion in the USA, and €10 billion within the European Union despite the fact that 91% of apps were free in 2013 (Cipriani *et al*., 2014; Bilyayeva *et al*., 2012).

App development mobile User Interface (UI) design is essential in the creation of mobile apps considering a range of factors including constraints, contexts, screen, input, and mobility as outlines for design. When developing medical mobile apps the content of the app should be based upon information that is useful to patients, and is evidence-based. The software release cycle for apps includes a series of stages from initial development to its eventual release:

- Pre-alpha: all activities performed during the software development but prior to testing;

- Alpha: the first phase of software testing using white box techniques (Khan, 2011);
- Beta: the phase generally begins when the software is feature complete, and generally incorporates usability testing to address speed and performance issues (Nielsen and Yoffie, 1994; Cusumano, 1998);
- Open and closed beta: closed beta versions of software have restricted accessibility based on the decisions of the developers. Open beta versions, by contrast, are tested by a much wider, informal group who are invited to report bugs (Apple, 2000; Microsoft, 2005);
- Full release of stable version.

Digital access

Innovation can sometimes be seen as a double-edged sword: when access is increased to a wider population it can sometimes limit access to other members of the population. This has been a noted concern about the growth of technology in healthcare, and the increased assumption that everyone has access to some form of electronic communication facility (Longley and Singleton, 2008). In 2010, 73% of households had internet access, compared with only 57% in 2006, and 73% of adults used the internet at least weekly, relative to just 51% in 2006 (Ofcom research report, 2013).

Technological developments for PROMs' routine data collection

Despite concerns about the ability to access and use electronic data capture systems in healthcare, there is a growing body of evidence which highlights the different systems in use, and has gathered feedback from users of all ages and in different health states. A range of different devices have been investigated. Boissy *et al.,* 2006 investigated bar-code scanning as a means of data entry on Personal Digital Assistants (PDAs) when compared to pen-and-paper to allow completion of self-report questionnaires. They identified that while participants found the system enjoyable to use and easy to access, there were concerns about the responsiveness of the system which could hinder wider-scale use. Dale and Hagen used PDAs in comparison with patient-completed diaries and found that while the PDAs performed better than pen-and-paper in terms of improved patient

compliance, increased data accuracy, reduced data handling time, and patient preference; technical malfunction was the biggest hindrance to use (Dale and Hagen, 2007). The issue of physical capability for transferring data has been highlighted in work by Russell *et al*., 2002. This study compared Internet and paper-based data collection but also investigated the use of two bandwidths *i.e.* ISDN at 128kbit and 17kbit. Bandwidth was found to have no significance on any of the measures suggesting that even home-based bandwidth provision could allow data to be submitted by patients at their place of residence if Internet access is available.

Other mobile devices have been tested including touch screen computer systems (Greenwood *et al*., 2006; Salaffi *et al*., 2009); this technology was found to be a perfectly acceptable option and was not affected by previous experience of computer use. Tablet computers have been tested in some settings. Horng *et al*. studied their use in an Emergency Department (Horng *et al.,* 2012). This was found to be a feasible option in a busy clinical setting; it was also found to be associated with a reduction in the amount of time clinicians had to log on to a computer. Horng *et al*. make the suggestion that this reduction in time at workstation computers could result in increased availability for patient contact, but this association will require further research to support or refute it (Horng *et al.,* 2012).

The global market in mobile smartphones, and applications continues to grow at a startling rate. Gartner has predicted that by 2016 there will be 310 billion apps which have been downloaded, and nearly two million in general distribution (Gartner, 2016). Apps are available in a variety of sites including Google play (Android apps), and Apple's app store (iOS or iPhone Operating System apps). The availability of apps for each operating system is shown in Figure 5.1 based on data from Statista.com, 2016.

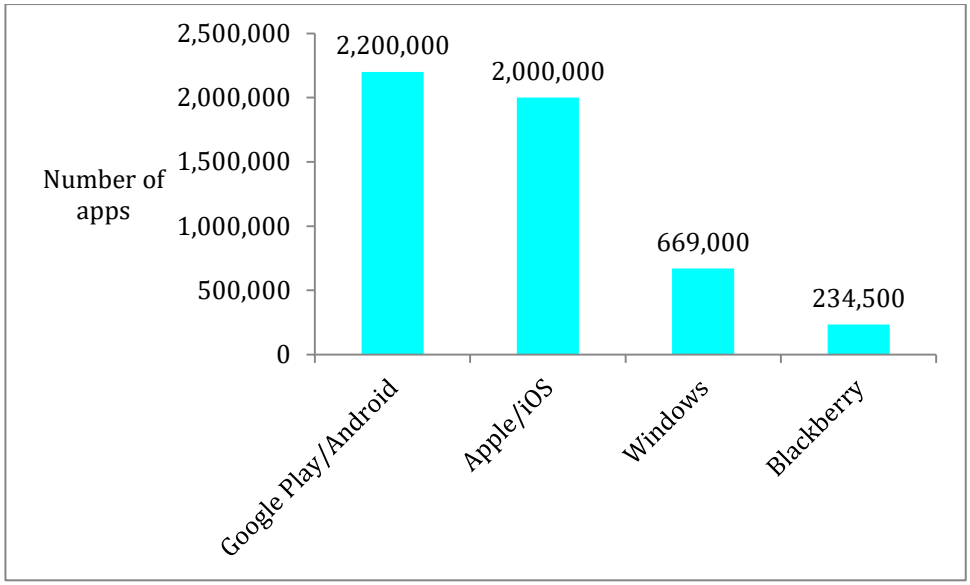


Figure 5.1 Estimated number of apps available for each operating system

(Data source: Statista.com (June, 2016)

The availability of apps may be a feature of smartphone sales: Gartner has analysed also smartphone sales to end users, and a graph based on their data is shown in Figure 5.2.

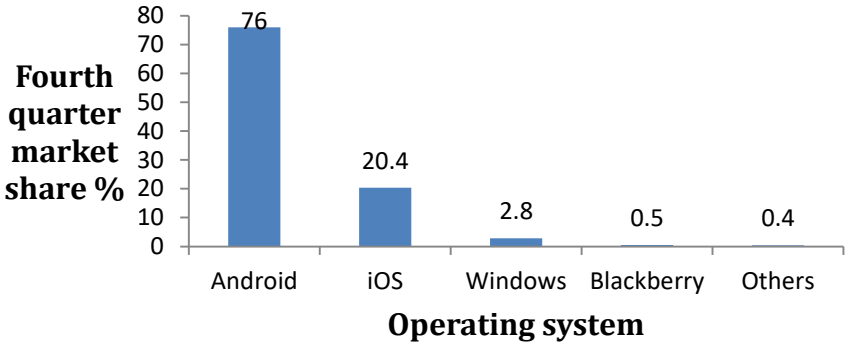


Figure 5.2 Worldwide smartphone sales to end users by operating system

(Data source: Gartner, 2016)

While the focus of access of many technology analytics is smartphones, laptops, desktops, and tablets remain a source of access to technology for a sizeable sector of the population. The manner in which this access has changed (2012-2015) and is predicted to change (2016) is shown in Figure 5.3. Although the technology market is clearly changing, the percentages indicate the importance of developing a PROM data collection system which is accessible for individuals who accessed the

Internet using a personal computer (PC), as well as being accessible via a tablet or smartphone app.

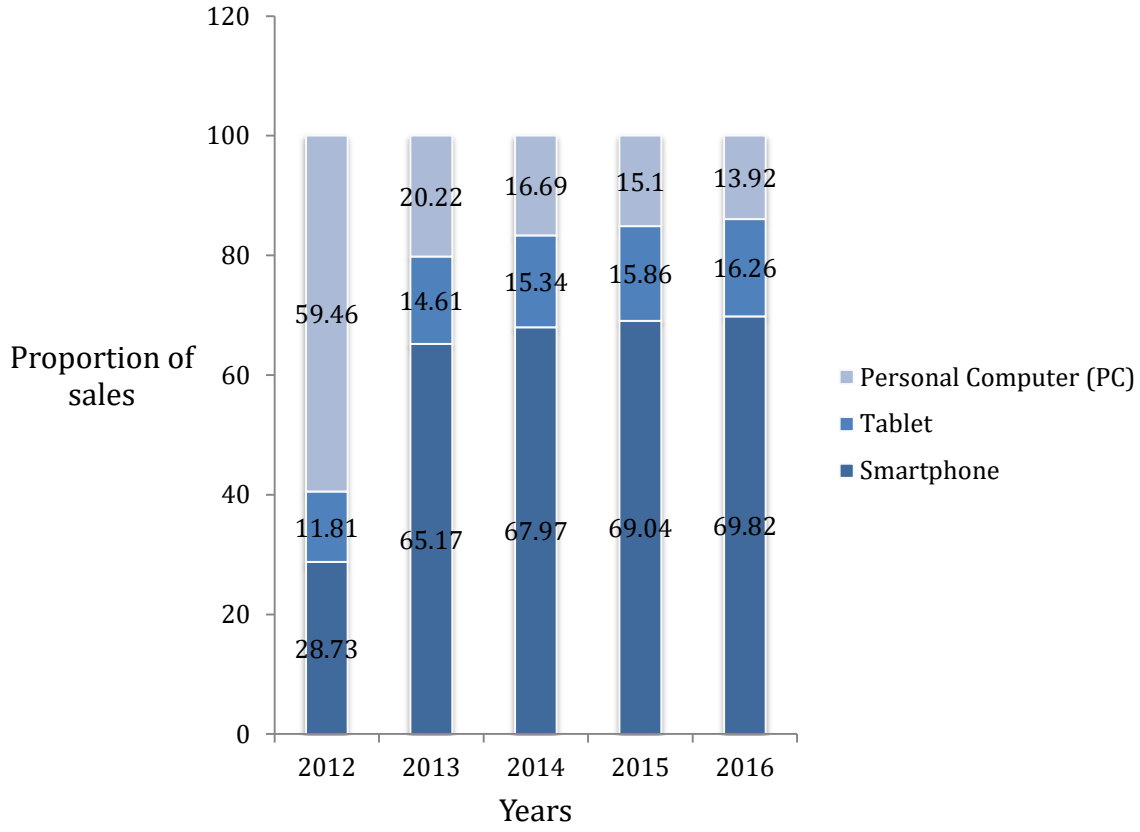


Figure 5.3. Actual and predicted access to technology devices 2012-2016. (Source: Titcomb, 2015, based on data from ICD Research).

Once the decision to develop the content for a web (online) app, and smartphone/tablet app had been made for this PhD, concern was given to other technological issues. A range of other key issues had to be considered but one significant one was the actual operating system in which the app would be developed, *i.e.* either Android or iOS (for Apple devices). Android and iOS are two competing operating systems continually developing to increase their market share. How an app is designed and interacts will be quite different between the two platforms. One of the main reasons for launching on Android is to reach a larger audience. The Android operating system was expected to be running on more than 1.9billion devices by the end of 2014, compared to about 700million for Apple's iOS operating systems. This is supported by the data in Figure 5.2 relating to the operating systems in smartphones purchased.

However, once developed, an app needs to be tested on the devices that will run it. In the case of iOS this is the iPhone and iPad. Apple builds the devices and the operating software which runs the system; more thorough testing involves the use of older devices also. In contrast, Android was developed as open source and is potentially more challenging to get right. The open source code has been adopted by different companies to run their smartphones creating a larger pool of handsets on which to test the app. In 2012 there were 4000 unique devices running Android rising to 12,000 by 2013. Approximately 600 different companies manufactured those devices. This apparently insurmountable task can be ameliorated by identifying the top 10 Android handsets and testing the app on these. Another consideration of the app development is the versions of the operating systems the app supports. Both Android and iOS have a substantial update about once a year. To make this process easier to accomplish, Apple revamped its operating system in late 2011 to allow updates to be made over the air. For the users of Apple devices, approximately 75% (approximately 230 million) were running the new 2011 operating system version. When considering the 600 different companies using Android devices, this means that up to eight different versions of Android can be running at any one time even though in practice the latest three versions account for the vast majority of devices being used. This situation has been improved now due to defragmentation of android operating systems and optimisation of apps by other handsets.

Such basic considerations are implicit in project management, but equally the need to adapt to a growing market for apps, meet expectations from patients about collecting feedback, and meet the demands of professional stakeholders and third party payers are also important. All of these considerations had to be evaluated against a background of creating a high quality product in as cost-effective a manner as possible. It is hardly surprising that this dilemma is referred to as the Devil's Triangle (Figure 5.4). The need to develop content for and commission the building of an app wisely, frugally, without wasting time, without excessive costs, and maintaining rigour and energy was not to be underestimated (Ries, 2011).

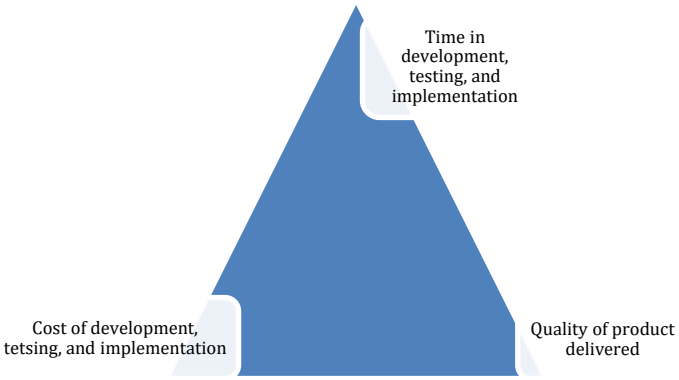


Figure 5.4. The Devil's Triangle of the app project management

Key decisions were taken about the PROM app based on these considerations:

- The app would be for mobile/tablet use;
- The app would be accessible online (web app);
- It would be developed for Android operating systems initially and tested on Android devices;
- The content of the app would be informed by the literature, and the output of the qualitative work involving patients and clinicians (Chapters 2 and 3), and the review of measurement properties (Chapter 4);
- The app content would be pilot-tested with a small number of patients for usability;
- The app would be field-tested with patients recruited by volunteer osteopaths in private practices, and patients in osteopathic educational institutions (OEIs);
- The app would be available in a refined form post-pilot by summer of 2015.

Once these decisions had been made, they were actioned starting with the commissioning of Clinvivo.com who would develop the app based on content supplied from this PhD. This app development was based on a Software as a Service (SaaS) model. A staged process was followed as described in Figure 5.5.

- Define the app objective;
- Design the content;
- Development of the app by Clinvivo.com;
- Test the app;

- Disseminate the app into clinical practice.

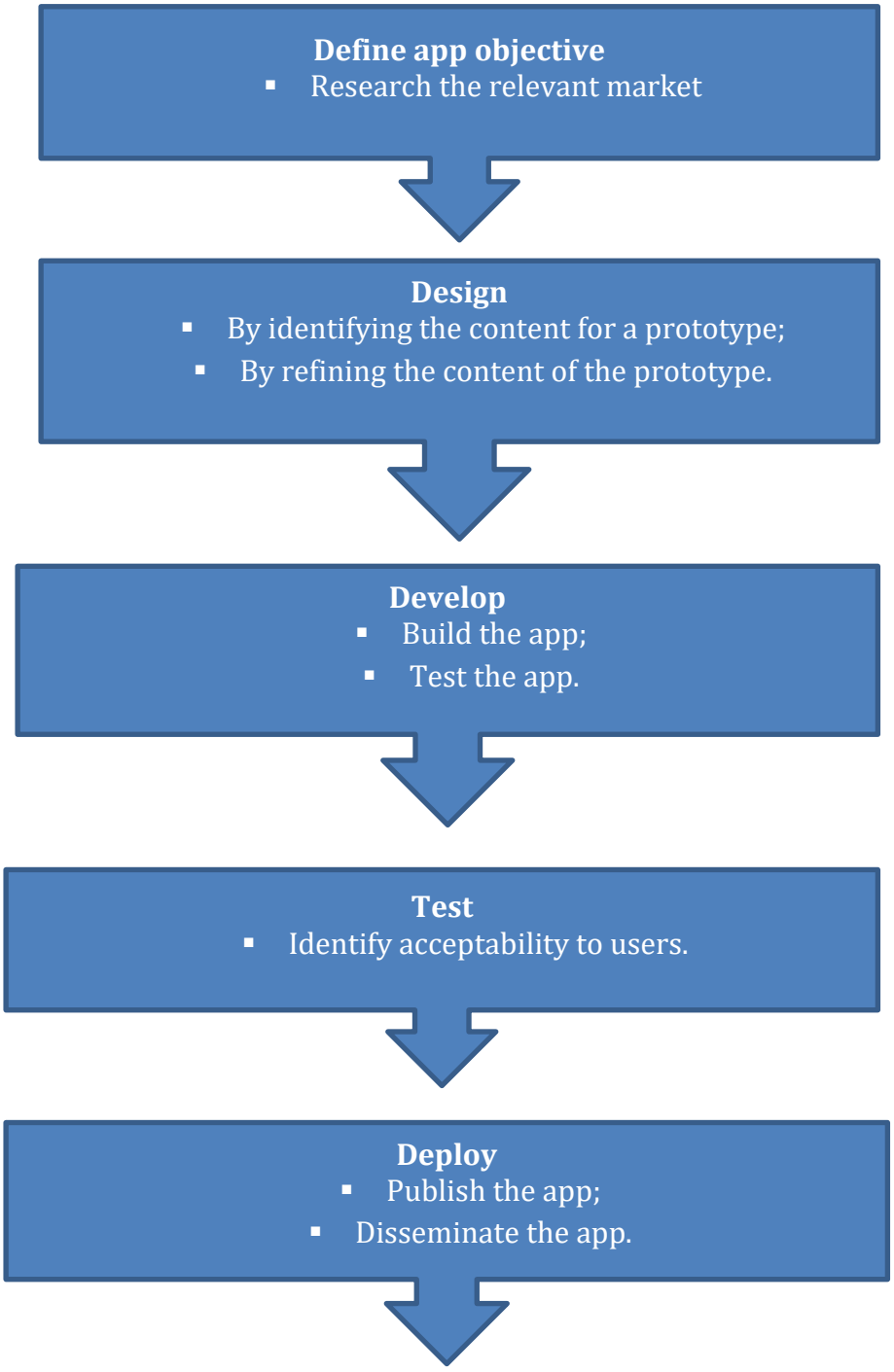


Figure 5.5 Developmental stages in the app development process.

Clinvivo's initial development featured several processes including identification of what other apps are available and doing the same or similar things; what design looks attractive to users; what technical information is available about competitor apps; and what potential opportunities are available to market the app, and generate revenue from its use.

Since osteopathy is considered to be a "manual therapy" profession, one of the obvious starting places was to contact colleagues in other manual therapy professions, namely physiotherapy and chiropractic to identify what initiatives they had introduced for electronic outcome data collection. The physiotherapy profession, through funding made available by the Private Physiotherapy Educational Foundation (PPEF), had developed an electronic standardised data collection (SDC) tool (Moore *et al*., 2012). This was available through a website but required referral to a code book to input information about presentation of symptoms, treatment modalities, and outcomes. This system has continued in use since its initial development. The chiropractic profession, in contrast, had already made significant progress in the area of practice-based outcome data collection through the Care Response system which continues to grow and is very successful. This facility functions as a free and pragmatic system to help practices gather and report clinical outcome and patient satisfaction information. It is promoted and supported by the Royal College of Chiropractors.

In this system, patients complete a questionnaire prior to treatment, and are emailed follow up questionnaires at 14, 30, and 90 days to allow assessment of change during the treatment process, and the long term effects of chiropractic management (Care Response; Newell *et al*., 2016).

Electronic data capture including PROMs is a growing market in healthcare. In response to this a series of companies have evolved, or explored new markets from existing data capture services. Contact was made with various companies to identify what they could provide, and the long term potential for system development, and their initial set-up and ongoing support costs (amplitude-clinical.com; clinvivo.com; documentcapture.co.uk; Fr3proms.com; quality-

health.co.uk; RioMed.com). At initial examination, some companies were clearly focussed on organisations like the NHS and clinical trials organisations with appropriate development budgets. Some mid-sized companies were able to provide examples of a bespoke system, but the ongoing support and scope for development were limited. Ultimately, Clinvivo Ltd was approached to work on the app development. This is a spin-off company based at Warwick Medical School with a qualified osteopath and senior researcher in outcome measurement as one of its directors. Clinvivo have an established record in data collection for clinical trials, and their expertise in this area was invaluable. Once this decision had been made, the focus was on clarifying the purpose of the app, and its associated content.

5.2 Purpose of the app

On examination of the various providers of e-data capture systems there was significant similarities in terms of data collected, and the manner in which feedback was provided. One significant feature of the Care Response system is the ability of clinicians to examine patients' responses to questionnaires while treatment is ongoing. The value of clinicians examining such feedback is much-debated in the literature, and the effect such examination might have on responses was considered carefully in the app development (Landsberger, 1958; McCambridge *et al*., 2014; Boyce *et al*., 2016). For this reason, the decision was taken not to provide feedback on individual patient data to clinicians.

One guiding feature concerning the purpose of the app for osteopaths was reflected in the initial application for the doctoral programme at QMUL. A series of aims for the PhD were identified, and these contributed to the development of the content for the app, and the ultimate data it could deliver. The purpose of the app development includes:

- Collect demographic data from osteopathic practice;
- Collect service data from osteopathic practice;
- Collect outcome data from osteopathic practice (initially focussing on musculoskeletal conditions);

- Collect governance data from osteopathic practice;
- Collate data in a secure environment;
- Develop content for data capture that functions well and without disruption to other software;
- Develop content for data capture that can be enhanced over time.

The content of the app must reflect information which is pertinent to the management of low back pain: this information includes factors relating to the prediction of outcome. Historically back pain has been regarded as a largely self-limiting condition with a favourable outcome after a defined period of time (Indahl *et al*., 1995; Malmivaara *et al.,* 1995). However, if the natural history of low back pain is examined over a patient's lifetime a rather different representation emerges (Burton *et al.,* 2004). Low back pain is reported as a recurring theme with varying levels of pain and disability throughout the lives of different individuals in a range of settings, work roles, and accompanying health states. The unpredictable nature of low back pain may, it has been suggested, be reflected in patterns of care seeking behaviour, but increasing numbers of studies have investigated not only the factors affecting the onset of low back pain, but those affecting outcome.

A range of factors have been found to affect outcomes in patients reporting low back pain. These include demographic factors, psychological symptoms, psychosocial factors, a patient's clinical history, and factors associated with employment (Harms *et al*., 2010). Although the literature examining the predictors of outcome for low back pain is growing, it has been criticised for the heterogeneity of populations, treatment settings, the presence of comorbidities, the different measures of outcome used, and the varying nature of time points at which follow up has been measured (Pengel *et al*., 2003). When considering different outcomes, the management of back pain involving invasive and non-invasive approaches has some overlap. Herbert *et al*. and Cook *et al*. studied patients with lumbar disc injury undergoing surgical management (Cook *et al*., 2015; Herbert *et al.*, 2016). They identified that although there was some variation depending on the PROM used, higher levels of pain at baseline, older patients, lower reports of disability, and higher quality of life scores were associated with

improved outcomes. This summary will focus now on the use of non-invasive approaches to the management of low back pain (LBP).

Outcomes have been examined among a range of different patient groups and exploring an array of factors. In examining patients experiencing low back pain during pregnancy, researchers have identified contrasting findings. Mogren identified that earlier onset of pain during pregnancy, and higher levels of pain were strong predictors of post-partum low back pain, while higher body mass index (BMI), higher maternal age, and a higher proportion of joint hypermobility were weaker predictors of lasting LBP (Mogren, 2006). More recent work by Peterson *et al.* failed to identify a single variable to be predictive of post-partum LBP (Peterson *et al*., 2014).

When considering sociodemographic factors, Moffett et al., identified that as Townsend scores increased indicating greater deprivation, levels functional disability increased also (Moffett *et al.,* 2009): being out of work also resulted in an increased disability score over time. Psychologically demanding work, and a poor expectation of being able to return to work were identified as predictors of poor outcome by Wippert *et al*. and Karsten *et al*. respectively (Karstens *et al*., 2013; Wippert *et al*., 2017). The presence of severe clinical stress, and the occurrence of critical life events were identified by Nordemann *et al*. and Wippert *et al*. respectively as being predictive of a poor outcome with activity limitation persisting after two years in women, but the presence of stress at work and its detrimental effect on outcome was supported by Karstens *et al*. and Lønnberg *et al*. but disputed by Ang *et al*. (Ang et al., 2008; Lønnberg *et al*. 2010; Karstens *et al*., 2013; Wippert *et al*., 2017).

Patients' beliefs about their pain and other symptoms have been examined. Negative illness beliefs were found to be a significant predictor of poor outcome (Ang *et al*., 2010; Glattacker *et al*., 2013). In other patients, poor self-efficacy and engagement in activity, increased number of care visits (to a GP or other clinician), and engaging in fear-avoidance activities were all found to be suggestive of poor

outcome (Iles *et al.,* 2008; Rasmussen-Barr *et al*., 2012; Karstens *et al*., 2013; Roberts *et al.,* 2015; Wippert *et al*., 2017).

Patients' expectations and satisfaction may be regarded as being another facet of illness beliefs. There is a perception that a good outcome is significantly associated with satisfaction. However, the relationship between satisfaction and outcome has been shown to be contradictory. In osteopathic research Pincus *et al*. found that outcome and satisfaction were poorly related, while Licciardone *et al*. found that there was a significant association Pincus *et al*., 2000; Licciardone *et al*., 2002). In chiropractic patients, Breen and Breen noted that change scores can have a weak to moderate impact on satisfaction (Breen and Breen, 2003). There has been a widespread belief for many years that positive patient expectation could influence treatment benefits. This has been a major rationale for the use of masking among clinical trial participants to their assigned treatment (Preston *et al.,* 2000). In studies examining the effects of patients' expectations, the findings have been contradictory. Myers *et al*. and Kaptchuk *et al*. identified that higher expectation, often mediated through an enhanced patient-clinician relationship, had a favourable effect on outcome (Kaptchuk *et al.,* 2008; Myers *et al*., 2008). In contrast, Sherman *et al.* found that pre-treatment expectation was not predictive of outcome (Sherman *et al.,* 2010).

The duration of symptoms and their intensity have been examined for their effect on outcome. Higher levels of pain have been suggested as being indicative of poor outcome (Rasmussen-Barr *et al.,* 2012; Glattacker *et al.,* 2013; Wippert *et al*., 2017), while longer duration of a compliant was suggested as indicating a poorer long term outcome by Karsten *et al*. (Karstens *et al*., 2013). High levels of pain over a prolonged period of time may result in increased levels of disability and subsequent impairment in daily life: both symptoms were suggested to predict poor outcome by Rasmussen-Barr *et al*., Karstens *et al*., and Nordemann *et al*. (Rasmussen-Barr *et al.,* 2012; Karstens *et al*., 2013; Nordemann *et al*., 2017).

There are considerable numbers of factors that have been proposed as influencing outcome, but other researchers have suggested that it is important to examine the

time at which such an assessment is made to be able to predict outcome more accurately. Vavrek *et al.* suggested that prediction of outcome should be made at six weeks post-treatment (Vavrek *et al*., 2015); other research has focussed on tools to support making more accurate assessments of outcome, and the recommendation for repeated use of such tools during a patient's management programme *e.g.* the STaRT Back Tool (Hay *et al*., 2008; Beneciuk *et al*., 2014). For other clinicians, the use of technological assessments *e.g.* Magnetic Resonance Imaging (MRI) should be used to evaluate outcome but research has shown that identification of spinal anomalies by MRI is a poor predictor of outcome (McNee *et al.,* 2011).

Predicting patient outcome clearly rests on a range of disparate factors. It is important for clinicians to be aware of such factors, and be able to manage those factors where appropriate through education, encouraging self-management, promoting general self-efficacy, or signposting to other clinicians where appropriate for support with psychological issues, pain management, or work-related issues.

Mindful of the competing demands discussed earlier of delivering a high quality app in a timely and cost-effective manner, initial scoping of content was undertaken. Initial ideas were included, discussed within the project team and with Clinvivo.com, and either accepted or rejected. The list of items discussed for baseline data collection is shown in Table 5.1.

Although treatment effectiveness is one of the most commonly used outcomes of care, other outcomes have equal significance, most notably where adverse or unexpected treatment reactions occur. This area of practice, in common with many others in healthcare, has suffered from a paucity of information in the past. Although the pharmaceutical industry implements pharmacovigilance and the yellow card scheme to detect adverse drug reactions in the wider population, other healthcare interventions have fewer structures in place where safety is concerned (Waller and Evans, 2002; MHRA, 2014). To address this lack of knowledge, and mindful of its necessity to support the consent process in practice, a series of

studies were commissioned by the osteopathic profession's regulator, the General Osteopathic Council (GOsC). These studies looked specifically at the risks associated with all manual therapies, how adverse events were defined, the most effective strategies for communicating risk to a wide cross-section of patients, the types of complaints made about treatment and/or osteopaths by patients to the regulator or insurers, and the risks associated with osteopathic management of patients (Carnes *et al*., 2009; Carnes *et al*., 2010; Leach *et al*., 2007; Leach *et al*., 2008; Vogel *et al*., 2013). In addition, studies were conducted in the teaching clinics of osteopathic educational institutions (OEIs), and reviews were conducted focusing on specific techniques or symptom presentations (Froud *et al*., 2008; Rajendran *et al*., 2009; Oliphant, 2004; Lisi *et al*., 2005; Gibbons and Tehan, 2006; Snelling, 2006). The draft content of the app was mapped onto a series of PowerPoint slides to aid discussion with Clinvivo.com, the app developers.

5.3 Selection and formulation of content

Discussion of the content for the pilot version of the app resulted in the following items for:

Baseline data collection.	
Demographic information	Age Sex Work status Ethnic background
Personal evaluation of general health status.	
Waiting time for the first appointment offered	
Duration of current symptoms	
Description of main area of symptoms (selection of multiple items from list)	
Main complaint/reason for seeking treatment	
Evaluation of severity of pain (marked on a VAS)	
Bournemouth Questionnaire	
Roland Morris Disability Questionnaire (24-item version)	

Free text box for any additional comments

Follow-up data collection

Satisfaction with osteopathic care

Experience of osteopathic care

Evaluation of global change

Evaluation of severity of pain (marked on a VAS)

Bournemouth Questionnaire

Roland Morris Disability Questionnaire (24-item version)

Free text box for any additional comments

All questions were compulsory with the exception of the question exploring ethnicity. Previous work on data collection with osteopathic patients identified that some patients from ethnic minorities found this question offensive (Fawkes *et al*., 2009). When completing any questionnaire, participants may wonder why researchers have chosen to ask or use particular questions or scales. This will be explored in the next section looking initially at the choice of a PROM in preference to other newer innovations in measuring outcome.

5.3.1 Why PROM and not PCOM or POEM?

Although debate continues concerning the value of PROMs to the clinical setting, they appear to be required to increasing degrees within various healthcare settings. This has led to the morphing of PROMs into different variations including Patient Centred Outcome Measures (PCOMs) which were used in the early 2000s (Hegarty *et al*., 2002; McGrath *et al*., 2003; NHS England, 2014), and a hybrid known as a Patient Outcome and Experience Measures or POEMs (Somner *et al.,* 2012).

PCOMs were described in the literature within the specialist area of oral medicine. More recently, NHS England has focussed on the development of PCOMs as a means of giving patients and carers a voice in determining the most valuable goals in the therapeutic process. This is especially pertinent when patients experience a range of health conditions and it is important for their views to be valued more

Table 5.1. Initial ideas for data to be collected at baseline.			
Item	Accept	Reject	Comments/rationale
Age	√		Adults ages from 18 onwards in bands
Sex	√		Male/Female
Work status	√		Simplified list, and ability to select multiple options e.g. if both a parent/carer and part-time worker
Ethnicity	√		Streamlined list of headings. Optional question as this has caused offence in previous data collection exercises.
Information on patients' comorbidities		√	Potentially superfluous data.
Medication used		√	Issues concerning recall bias, and difficulties discriminating between prescribed and self-prescribed over the counter medication.
Duration of current symptoms	√		Useful data to identify numbers of patients in acute, sub-acute, and chronic symptom states.
Main reason for seeking treatment	√		To identify patients seeking treatment, and those patients requiring second opinion or advice only.
Areas of symptoms	√		To identify number of patients with single and multiple site symptoms.
Measurement of pain status	√		To identify baseline pain in patients seeking treatment, and subsequent change in pain level.
Outcome Measure	√		To measure outcome of care when different pertinent items are considered.
Comment box	√		Area for patients to add free text to clarify information given or comment on questions asked.

Item	Accept	Reject	Reason for rejection/comments/rationale
Measurement of pain status	√		Ability to capture change in pain scores pre- and post-treatment.
Outcome measure	√		To measure outcome of care when different pertinent items are considered. Change in outcome measurement scores will be calculated.
Medication change		√	Difficulty of obtaining accurate data, and lack of clarity concerning the value this data would add.
Patient satisfaction measure	√		To measure patients' level of satisfaction with their care.
Patient experience measure	√		To measure patients' experience of osteopathic care.
Global change measure	√		To measure overall change in symptoms at the time of completion of the follow up (1 week or 6 weeks post treatment).
Adverse events data		√	Concern about the challenge of collecting this type of data in a meaningful and accurate manner. Concerns about providing symptoms which might lead to speculative information.

highly than clinicians deciding what is best for them. (https://www.england.nhs.uk/ourwork/pe/pcoms/). In 2015, the focus for NHS England has been the development of PCOMs for children and young people. Seven sites across the UK (London (2); Liverpool (1); Bristol (1); Nottingham (1); Shropshire (1), and Stockton-on-Tees (1)) will be working with young patients and their families to develop measures for a range of health conditions including asthma, complex respiratory conditions, self-harming, eating disorders, palliative care, and for the users of wheelchair and posture services (NHS England, 2015 https://www.england.nhs.uk/2015/02/pcoms-cyp/). Although the revised versions of PCOMs sound very attractive with renewing focus on the role of the patient in healthcare, they lack current availability for use in day-to-day clinical practice.

PROMs rather than POEMS or PCOMs were selected for inclusion in the pilot version of the app: the PROMs identified were the 24-item version of the Roland Morris Disability Questionnaire (RMDQ), the Bournemouth Questionnaire, and a Visual Analogue Scale (VAS). Their inclusion was based on patient input as described in Chapter 2. The Bournemouth Questionnaire was described in detail in Chapter 4; the RMDQ and the VAS are described in the next section.

5.3.2 The Roland Morris Disability Questionnaire (RMDQ)

The original version of this Patient Reported Outcome Measure (PROM) is composed of 24 items measuring 12 separate categories which cover pain intensity, self-care, social life, walking, sitting, standing, sleeping, bending, stairs, appetite, general activity, and the ability to manage household chores (Roland and Morris, 1983a; Roland and Morris, 1983b).

Scoring and interpretation. Patients or research participants are asked to select items relevant to them, which score one point to questions that begin "because of my back". The maximum score is 24, and the minimum score is zero. The higher the score, the greater the disability experienced by the patient. Modified versions of the RMDQ exist which vary in the number of questions they contain, their content, and their response options (Roland and Morris, 1983; Patrick *et al*., 1995;

Stratford and Binkley, 1997; Underwood *et al*., 1999; Williams *et al*., 2001; Atlas *et al.,* 2003; Stroud *et al*., 2004; Kent *et al*., 2015), and one 11-item version of the 23-item RMDQ (Cook *et al*., 2008). The content of the different versions is shown in Table 5.2. Information concerning each of the different versions will be described in brief in the next section.

RMDQ-23

The RMDQ-23 was created by Patrick *et al.* by removing five of the questions in the RMDQ-24, and adding four items from the Sickness Impact Profile (Bergner *et al*., 1981: Patrick *et al*., 1995). It contains 23 items which measure activity limitation due to back and leg pain "today". It was regarded by Patrick *et al.* as having increased responsiveness, as it is able to measure pain that is both back and leg-related. Scoring is dichotomous.

RMDQ-18

The RMDQ-18 is a shorter modified version of the RMDQ-24, and two separate versions have been developed (Stratford and Binkley, 1997; Williams *et al*., 2001). In the version modified by Stratford and Binkley, questions 2 (changing position), 15 (appetite), 17 (walking), 19 (help with dressing), 20 (sitting due to pain), and 24 (remaining in bed) have been removed. In the version modified by Williams *et al.,* questions, 2, 15, 19, 20, and 24 have been removed while 17 remains, and 22 (irritation and bad temper) has been removed.

RMDQ-16

The RMDQ-16 is a modified form of the RMDQ designed to measure functional limitations during the past two weeks, and patients can respond with either "yes", "no", "don't know" or "not applicable". The wording is expressed slightly differently *e.g*. "In the past two weeks, because of past or present back pain have you…..

1. Stayed in bed more?
2. Done less of the jobs you would normally do around the house? *etc.* The full list of included questions is shown in Table 5.3. Scoring is achieved by:

$$\frac{\text{Number of affirmative answers}}{\text{Number of questions answered}} \times 100$$

The final score is expressed as a percentage with higher scores representing greater limitations to the activities of daily living (Longo *et al*., 2010).

RDQ-12

The RDQ-12 is also known as the Maine-Seattle Back Questionnaire. It is a 12-item version of the RD-23 questionnaire. The final score is obtained from the sum of all responses producing a range of scores from 0 (no impairment) to 12 (severe impairment).

RDQ-11

Two versions of the RDQ-11 have been produced. One version by Stroud *et al.,* 2004 is an 11-item version of the 24-item RMDQ, and the other modification is an 11-point version of the RMDQ-23 (Cook *et al.,* 2008). In the version by Stroud *et al*. Questions 1, 2 ,8, 13,14,15,18, 19, 20, 22, and 24 have been removed. Davidson, 2009 noted that while all of the authors of the revised versions have argued that their short-form versions are comparable in providing patient information, the loss of information appears to be important to patient management or research findings and could produce increased ceiling effects. The time for completion of the 24-item questionnaire has been estimated to be five minutes (Longo *et al*., 2009) so using shortened versions might appear futile.

Table 5.2 Contents of the different English language versions of the RMDQ

24 item RMDQ	23 item version (Patrick *et al*., 1995)	18-item version (Stratford and Binkley, 1997)	18-item version (Williams *et al.,* 2001)	16-item version Dionne *et al.*, 1997	12-item version (Atlas *et al.,* 2003)	11-item version (Stroud *et al*., 2004)	11-item version (Cook *et al.,* 2008)	RDQ-two (Underwood *et al*., 1999
1. I stay at home most of the time because of my back.	√	√	√	X		X		√
2. I change position frequently to try and get my back comfortable.	√	X	X	X		X		√
3. I walk more slowly because of my back.	√	√	√	X		√		√
4. Because of my back, I am not doing any of the jobs I usually do around the house.	√	√	√	√ (with modified wording)		X		√
5. Because of my back, I use a handrail to get upstairs.	√	√	√			√		√
6. Because of my back, I lie down to rest more often.	X	√	√	√ (with modified wording)				√
7. Because of my back, I have to hold onto something to get out of an easy chair.	√	√	√	X		√		√
8. Because of my back, I try to get other people to do things for me.	Modified – see end of table	√	√	√ (with modified wording)		X		√
9. I get dressed more slowly than usual because of my back.	√	√	√	√ (with modified		√		√

				wording)				
10. I only stand up for short periods of time because of my back.	√	√	√	X		√		√
11. Because of my back, I try not to bend or kneel down.	√	√	√	X		√		√
12. I find it difficult it get out of a chair because of my back.	√	√	√	X		√		√
13. My back is painful almost all the time.	X	√	√	X		X		√
14. I find it difficult to turn over in bed because of my back.	√	√	√	X		X		√
15. My appetite is not very good because of my back pain.	X	X	X	√ (with modified wording)		X		√
16. I have trouble putting on my socks (or stockings) because of the pain in my back.	√	√	√	X		√		√
17. I only walk short distances because of my back pain.	√	X	√	X		√		√
18. I sleep less well because of my back.	√	√	√	X		X		√
19. Because of my back pain, I get dressed with help form someone else.	X	X	X	X		X		√
20. I sit down for most of the day because of my back.	X	X	X	X		X		√
21. I avoid heavy jobs around the house because of my back.	√	√	√	√ (with modified wording)		√		√
22. Because of my back pain, I am more irritable and bad tempered with people than usual.	√	√	X	√ (with modified wording)		X		√
23. Because of my, I go upstairs more slowly	√	√	√	X		√		√

than usual.								
24. I stay in bed most of the time because of my back.	√	X	X	√ (with modified wording)		X		√
My back or leg is painful almost all the time.	√	X	X	X		X		
I have decreased sexual activity.	√	X	X	√ (with modified wording)		X		
I keep rubbing or holding areas of my body that hurt.	√	X	X	√ (with modified wording)		X		
I am doing less of daily work around the house.	√	X	X	√ (with modified wording)		X		
Express concern about my health.	√	X	X	X		X		
Done fewer social activities with groups of people?				√				
Talked less with those around you?				√				
Not kept your attention on any activity for very long?				√				
Not finished the things you start?				√				
Accomplished less than usual at work?				√				
Gone out for entertainment less often?				√				

RDQ-7p

The RDQ-7p is another modified version of the 24-item RMDQ. The layout uses a seven point Likert scale where "0" represents "disagree strongly", "3" represents "not sure", and "6" represents "agree totally".

	0	1	2	3	4	5	6
1. I stay at home most of the time because of my back.							

The final questionnaire score is expressed as:

$$\frac{\text{Total patient score}}{\text{Total possible score}} \times 100$$

Once again, the higher the score the greater the disability.

RDQ-two

The RDQ-two is a modified version of the RMDQ designed to assess low back pain over the preceding four weeks. Responses are expressed by:

1. I have stayed at home because of my back…*etc*.

Not at all	1-7 days	8-14 days	15-21 days	22-27 days	Every day	Not applicable

The number of the questions and the domains investigated are the same as the original 24-item version. Scores are awarded for each response.

Response	Score
Not at all	0
Not applicable	0
1-7 days	0.2 points
8-14 days	0.4 points
15-21 days	0.6 points
22-27 days	0.8 points
Every day	1 point

The final score is calculated by:

$$\frac{\text{Total patient score}}{24} \times 100$$

Although different versions of the RMDQ exist, it is the 24-item version that has undergone most evaluation. Its evaluation has identified that a patient is considered to have truly changed when the difference between the previous score and the current score exceeds the combined error associated with both measurements. The magnitude of this error is described as the minimal detectable change (MDC) (Stratford and Binkley, 1996; Stratford and Binkley, 1999). The MDC is also estimated from the standard error of measurement (SEM): conditional standard errors from both the previous score and the current score are applied to estimate the MDC (Stratford and Binkley, 1999). The SEM represents the within-patient variability which can occur when several measurements are obtained from the same patient (Stratford and Binkley, 1996; Stratford *et al*., 1996).

The original questionnaire and all translations are in the public domain. No permission is required for their use or reproduction. A total of 36 translations and adaptations are available, including minor adaptations for US, Canadian and Australian English (http://www.rmdq.org/). The translation process recommended is described in the work by Beaton *et al*., 2000, and the majority of this work is undertaken by MAPI.

5.3.3 Methods of measuring pain

Pain is one of the most common symptoms for which patients consult a medical practitioner, and an osteopath (Fawkes *et al*., 2010). As a consequence, pain assessments are often regarded as the fifth vital sign (JCAHO, 2013).

A number of methods exist to measure pain. They include self-report measures involving questionnaires devoted explicitly to pain, self-report measures embedded within questionnaires, and rating scales which can be applied to pain measurement alone or other symptom measurement also.

The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations while aimed primarily at improving clinical trial methodology have a relevance to clinical practice where pain is a feature for consulting patients. IMMPACT recommended six core domains when measuring pain including:

- Pain;
- Physical functioning;
- Emotional functioning;
- Patient rating of improvement, and satisfaction with treatment;
- Patients disposition, and characteristics;
- The presence of other symptoms, and adverse events during treatment (Dworkin *et al*., 2005; Dworkin *et al*., 2008).

The measurement of pain alone will be the focus of this section.

Defining pain in clinical practice

Chambers English Dictionary includes several definitions of "pain". They include "penalty; suffering; bodily suffering; great care or trouble in doing something; the throes of childbirth; a tiresome or annoying person; to cause suffering to; and to put to trouble" (Chambers, 2014). When the very definition is so complex, it is hardly surprising that one of those definitions is equally multi-faceted. There are, however, several aspects to consider when defining pain and its effects (Von Korff *et al*., 2000). They include:

- Pain severity;
- Pain chronicity;
- Pain experience (Haefeli and Elfering, 2006);
- Pain site;
- Factors affecting pain (Breivik *et al*., 2008).

Pain severity includes the intensity of the pain itself, and the level of interference to day-to-day activities that the pain produces (disability). While pain intensity can be quantified directly using a range of scales and instruments, some of which will be discussed later, pain-related disability measures are seen as a major indicator for the degree of pain severity (Haefeli and Elfering, 2006). To support this concept of using pain-related disability as a unitary construct, examination of different measures has indicated high levels of correlation (Kerns *et al*., 1985; Bergstrom *et al*., 1998).

Chronicity describes the enduring nature of pain experienced by patients. Different classifications of pain chronicity exist; Nachemson and Bigos defined it as a period of three months of persistent pain (Nachemson and Bigos, 1984). The International Association for the Study of Pain (IASP) defined pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage", and further defined chronic pain as "pain which has persisted beyond normal tissue healing time", taken, in absence of other criteria, to be 3 months" (IASP, 1986). Von Korff and Saunders reflected upon the more intermittent nature of pain, and defined chronic pain as that lasting for at least half of the days during a year (Von Korff and Saunders, 1996). In a more systematic examination of chronicity, Raspe *et al.* investigated the definitions of chronic pain in a total of 40 studies of low back pain. Chronic pain was defined as that lasting between 4 weeks and greater than 1 year (Raspe *et al*., 2003).

Pain experience includes the concept of pain intensity, and the degree of emotional arousal or changes in action readiness caused by sensory experience of pain (known as pain affect) (Von Korff *et al.,* 2000).

Many different measures exist to evaluate pain in clinical practice. Specific PROMs aimed at pain evaluation exist *e.g.* the McGill Pain Questionnaire is one of the most frequently used PROMs in clinical and research settings (Melzack, 1975; Rowbotham and MacIntyre, 2003). However, the most commonly used pain measures in clinical trials are the VAS and the NRS.

Visual Analogue Scale (VAS)

A VAS is a common form of response ubiquitous in outcome studies in all specialties, and commonly used to measure pain. It was initially published in the 1920s although it was not widely used at the time (Hayes and Patterson, 1921; Freyd, 1923; Aitken, 1969). Scott and Huskisson (1976) examined different pain scales in a range of experiments, and they identified that the VAS and graphical rating scale were the most sensitive and satisfactory of all of the scales when used

horizontally with words spread along the length of the line. A variety of different formats can exist in terms of line length, and the verbal anchors used. For example line lengths can be 100mm or 10cm; Hjermstad *et al.* identified a range of anchor labels as shown in Table 5.3 (Hjermstad *et al*., 2011).

A VAS is considered to reduce the confounding effects of the variations that can occur between individual interpretation of the grading used in rating scales (Kersten *et al.*, 2012). It is preferred by patients who regard their pain response is not represented by the gradations in a grading scale, and subsequently enables patients to provide a finer distinction (Aitken, 1969; Hjermstad *et al.,* 2011). Visual analogue scales which use adjectives within the scale are termed graphical rating scales (GRS).

An example of a VAS and GRS is shown in Figure 5.6. With the increase in technological developments and their use in healthcare, electronic versions of the VAS are being developed and their validity and reliability are examined later in this section (Jamison *et al*., 2002; Couper *et al*., 2006).

Numerical Rating Scale (NRS)

The NRS exists in a variety of formats *e.g.* a NRS-6, 7-, 10- 11-, 20-, 21- or 101-point scale. It has end points of "no pain" and "pain as bad as it could be" or "worst pain imaginable" at its extremes, and a patient is asked to rate his/her pain from 0-10 (11-point scale), or 0-100 (101 point scale)(Jensen *et al.,* 1986; Ekblom and Hansson, 1988; Herr and Mobily, 1993; Carpenter *et al*., 1995; Bergh *et al*., 2000; Svensson, 2000; Williams *et al.,* 2000; Bergh *et al*., 2001; Lundeberg, 2001; Singer *et al*., 2001; Herr *et al.*, 2007; Huber *et al*., 2007). The NRS has been shown to be valid and reliable, demonstrating significant correlations with other pain intensity measures *e.g.* VAS (Jensen *et al.,* 1986). Forms of the NRS are often embedded in other measures *e.g.* the Brief Pain Inventory (Cleland and Ryan, 1994). The NRS is straightforward to administer and score improving its utility for older patients or

Table 5.3 Variations in anchor labels in studies using a VAS

Anchor labels	Number of studies identified, n
No pain, worst pain	5
No pain, worst pain possible	3
No pain, the worst pain possible	8
No pain, worst pain imaginable	11
No pain, worst pain ever	3
No pain, pain cannot be worse	1
No pain at all, unbearable pain	5
No pain, pain as bad as it could be	4
No pain, very intense pain	1
No pain, the most intense pain imaginable	4
No pain at all, sever pain	3
No pain, the most severe pain you can possibly imagine	1
No pain, the most intense pain sensation imaginable	3
No pain, maximum pain	3
No pain, maximal amount of pain	1
Least possible pain, worst possible pain	3

others with motor impairments, and its compliance. The data are interval. An example of a NRS is shown in Figure 5.6.

Verbal Rating Scale (VRS)

The VRS consists of a list of adjectives describing different levels of pain intensity spanning from "no pain" to "extremely intense pain". Several different versions of the VRS exist: patients are requested to read the descriptions and select the word they feel describes their pain most appropriately (Joyce *et al.*, 1975; Seymour, 1982; Frank *et al.*, 1982). Adjectives are assigned numbers as a function of their

rank. The rank numbers can give the impression that the intervals between each description are equal, but this is not the case (Jensen and Karoly, 1992; Turk and Melzack, 2001). Statistical evaluation can be problematic as the data can be treated incorrectly: VRS data are ordinal. The VRS demonstrates good compliance (Jensen *et al.,* 1986; Jensen *et al.,* 1989), and demonstrates sensitivity to treatments which are used in the management of pain intensity (Fox and Melzack, 1976). An example of a VRS is shown in Figure 5.6.

Other measurement scales for pain exist including Picture or Face scales (Keck *et al*., 1996; Beyer and Knott, 1998). Patients are asked to indicate which of the included facial expressions showing pain best represents their experience of pain. These scales are particularly helpful in children and in adult populations (Wong and Baker, 1988; Stuppy, 1998). They have been shown to be valid when compared to other measures of pain intensity (Bieri *et al.,* 1990; Beyer and Knott, 1998).

Measurement properties of VAS, and NRS.

The extensive literature on the use of VAS and NRS identifies that overall scores for VAS and NRS correspond well although there are a few exceptions where systematically higher scores of VAS are recorded. Hjermstad *et al*. in their review of 19 studies identified the NRS had higher compliance rates and better responsiveness, and ease of use when used in a paper format (Hjermstad *et al.,* 2011). Overall they found that the VAS was the most frequently used scale: VAS was used 59 times in 52 of the 54 studies examined compared with 37 studies using a NRS. The format of the VAS varied between studies, but the more traditional horizontal version was favoured (N=55) to the vertical version (N=4). This change in orientation has been found to impact upon the distribution of scores: Ogon *et al*. found that data were normally distributed when used horizontally but not vertically (Ogon *et al*., 1996). Data obtained by horizontal and vertical use of VAS has been found to have low levels of agreement (Dixon, 1986), but correlates well (Scott and Huskisson, 1979, Hinchcliffe *et al*., 1985).

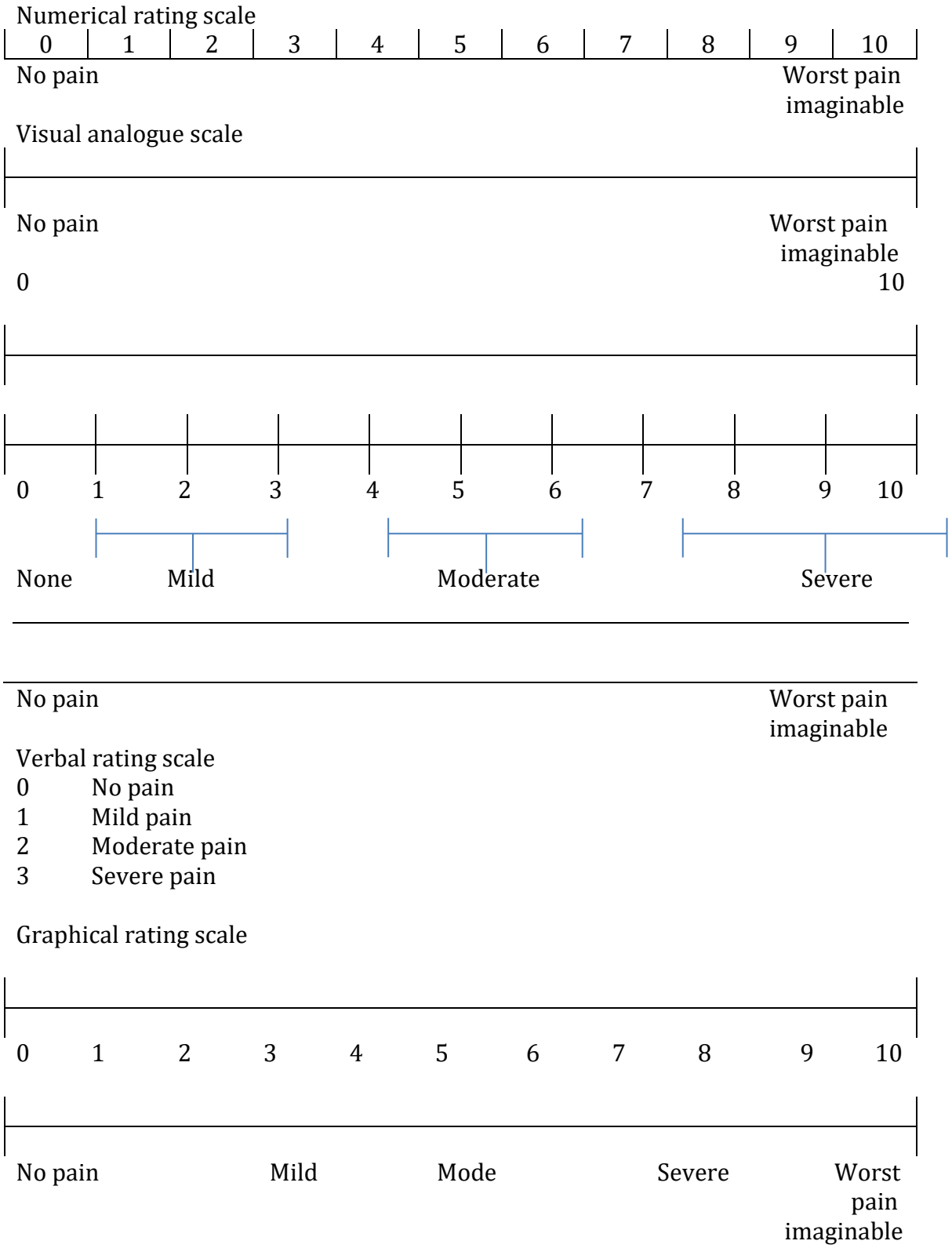


Figure 5.6 Common versions of pain scales

Electronic and paper versions of pain scales

Although the paper version of VAS is an established and validated method for measurement of pain, its use in electronic media is less well-tested. In addition, when considering electronic data capture, consideration must also be given to how the data are collected, for example, using either sliders (an electronic marker which can be moved vertically or horizontally to adjust a variable) or radio buttons (icons representing one of a set of option responses to a question, only one of which can be selected at any time). Recent technological developments have seen the incorporation of the VAS and other self-report measures of pain into palmtop computers, Tablet computers, and laptops (Jamison *et al*., 2002; Couper *et al.,* 2006). While the paper and electronic version of the VAS are similar in appearance, the electronic version of the VAS (eVAS) has the appearance of being shorter in overall length, in the ease with which to make a mark upon the line using a stylus or finger, and erase earlier marks if they were regarded by the patients as being inaccurate (Jamison *et al*., 2002). Although the eVAS has been used in clinical trials since 1995, establishing its responsiveness and reliability is important (Tiplady *et al.,* 1995). A range of literature has now examined this process. Jamison *et al.* undertook a crossover study of 24 healthy subjects using a 6cm square Palm Pilot IIIxe (Jamison *et al.,* 2002). The eVAS capture programme linearly converted the pixel touched into an integer between 0 and 100. Verbal and sensory methods of data capture were compared, and response recording methods (paper and electronic) were compared also within the trial of the four experimental conditions (N=2016) based on standards for psychophysical evaluation (Gracely, 1988; Gracely, 1989). Support was found for the eVAS version with multivariate analyses showing equivalent stimuli to be rated the same whether entered using eVAS or paper VAS.

In later work, Couper *et al.* compared the VAS when data were contributed using either radio buttons, numeric input or sliders (Couper *et al*., 2006). In this early experimental work they found that the VAS slider had higher rates of non-completion, higher rates of missing data, and longer completion times. New innovations in including sliders and other items in apps make them more sensitive to use, and the rates of missing data will be assessed as part of this study.

It is important to consider measurement properties for pain in a range of settings. Bijur *et al.* examined the test-retest reliability of VAS in the acute pain setting, and identified that 90% of the scores are close together when repeated within a short time frame (Bijur *et al*., 2001). Statistical values of reliability for the VAS have been calculated, and show good levels at test-retest with intra-class correlation coefficient (ICC) ranging from 0.99 (95%CI: 0.989-0.992) to 0.97 (95%CI:0.96-0.98) (Bijur *et al*., 2001 and Gallagher *et al*., 2002 respectively). More recent evaluation using an IPad in a healthy population has identified ICC values of 0.90 (95%CI: 0.82 – 0.95) (Bird *et al.,* 2016).

The NRS has been investigated less extensively than the VAS for reliability. Van Tubergen *et al.* assessed this in their study of 536 patients with ankylosing spondylitis but concluded that reliability was poor (Van Tubergen *et al*., 2002). However, they reported that "the huge differences in scores give the impression that some patients did not fully understand or properly read the anchors of the scales".

While measurement of modalities such as pain and disability are both important and widespread in the literature, increasingly clinical governance requires the measurement of patients' satisfaction and experience. These two properties will be considered in section 5.3.5.

5.3.4 Evaluating measurement properties

The reliability and responsiveness of PROMs are key considerations when using them for particular patient groups or in specific settings. These properties and the factors to consider in their interpretation will be considered in turn.

Reliability

Reliability is defined as "the degree to which an instrument is free from measurement error" (Mokkink *et al*., 2010a). However, an extended definition is highly pertinent to the use of PROMs in measuring change in patients and is "the extent to which the scores for patients who have not changed are the same for repeated measurement under several conditions" (Mokkink *et al*., 2010a).

Reliability refers to the consistency of the scores; it is a characteristic of an instrument used in a population and not just of an instrument.

Reliability parameters range in value from -1 (totally unreliable) to +1 (perfectly reliable). One of the most frequently used reliability parameters is the intra-class correlation coefficient or ICC (Shrout and Fleiss, 1979; McGraw and Wong, 1996). There are several ICC formulas and all consist of a ratio of variances; variances can be calculated using one way Analysis of Variance (ANOVA). The calculation of Pearson's *r* is cited also in measures of reliability but its values tend to be higher than those for ICC and it is regarded as being less stringent (de Vet *et al.,* 2009). Although Cronbach's α is sometimes used instead of ICC, it is recommended that it should not be used instead of ICC if measurement error is a statistic of interest (de Vet *et al*., 2011). Cronbach's α is considered to be at an acceptable when values exceed 0.7 but are not higher than 0.9 (Streiner and Norman, 1995). The value of Cronbach α is highly dependent on the number of items in the scale. A large number of items can give a high Cronbach α score even though inter-item correlation may be low.

When designing a study to evaluate the test- retest (TRT) reliability of a PROM it is important to bear in mind the situation for which the reliability is being measured (de Vet *et al*., 2011). The study sample should reflect the population of interest since reliability is highly dependent upon the characteristics of the study population. A range of time intervals are described in TRT studies *e.g.* from 2 hours to four weeks (Hartvigsen *et al*., 2005; Gay *et al.,* 2007). There are no standard rules for TRT time intervals and common sense is recommended when considering the stability of the patients' symptoms, any interference in symptoms from repeated evaluations, and the prospect of recall bias or panel conditioning of the time between administrations is too great or too small respectively (Mausner and Kramer, 1985; Underwood *et al.,* 2006).

When considering sample sizes for reliability studies, general guidance includes involving around 50 participants (Altman, 1991; EMGO, 2014). An adequate sample size is required to obtain an acceptable confidence interval (CI) around the

reliability parameter. Giraudeau and Mary, 2001 provide a formula based upon the intended ICC value, and the number of repeated measurements being taken. The more measurements being taken reduce the necessary sample size. This is an important logistical consideration in studies.

Reliability can be said to vary at group and individual level with higher values for reliability are required for measurements in individual patients. This is important since measurement for an individual may be followed by a specific therapeutic decision.

Responsiveness

The concept of responsiveness was discussed initially in the literature in the 1980s (Deyo and Centor, 1986; Guyatt *et al.,* 1987). Since that time there has been much discussion in the literature about whether responsiveness should be considered as a separate measurement property or whether it is in fact an aspect of validity *i.e.* the degree to which an instrument truly measures the construct it purports to measure". In essence responsiveness relates to the validity of a change score in contrast to a single score when considered at a single point in time.

Many different definitions of responsiveness can be found in the literature (Terwee *et al.,* 2003). While there are some definitions which are close in their definitions, others vary significantly. For example, some researchers regard responsiveness as "the ability to detect change in general" irrespective of whether this occurs due to true change, change in a different construct to that being evaluated, or due to "noise". Other researchers have defined responsiveness as "the ability to detect clinically important change" although this definition, by its very nature, must be accompanied by information concerning what constitutes a clinically important change. This raises an issue around the validity of the change score.

The assessment of a change in disease status in patients is an important aspect of patient management. While the attribute of responsiveness has been defined in many ways in the literature, I refer to the definition used by the COSMIN group as

"the ability of an instrument to detect change over time in the construct to be measured" (Terwee *et al*., 2003; Mokkink *et al.,* 2010a). Essentially, when measuring change in a patient's health status over time using a particular PROM, the patient's change in status will be reflected by a change in the PROM score. There have been major discussions in the literature concerning whether responsiveness should be regarded as a separate measurement property or whether it should be regarded as an aspect of validity (Deyo and Centor, 1986; de Vet *et al.,* 2011). In the context of this study, although I agree that responsiveness is an aspect of validity, I am looking at the validity of a change in scores longitudinally, and not at the validity of a single score.

Mindful of the requirement of measuring changes in scores longitudinally, at least two measurements are required to evaluate the responsiveness of an instrument. There are no recommended time periods for the measurement of responsiveness and much will depend on the condition being evaluated. Consequently, time periods between measures can range from a few weeks to a few months as long as a change can be anticipated within the interval selected (de Vet *et al*., 2011). There are various approaches to measuring the responsiveness of an instrument, but the two main approaches are the construct approach and the criterion approach.

The most frequent external criterion is a transition question which is a single question which asks a patient about their global perception of change to their health status whether improvement or deterioration (Jaeschke *et al*., 1989; Redelmeier *et al.,* 1993; Bombardier *et al*., 1994; Juniper *et al*., 1994; Barber *et al.,* 1996; Beurskens *et al*., 1996; Guyatt *et al*., 1998; Guyatt, 2000b).

In this study the criterion approach has been used to measure the responsiveness of the BQ, RMDQ, and VAS in an electronic format with a Transition Question being used as a "gold standard" against which to measure change. The Transition Question is discussed in greater detail in section 5.3.7. Although there is some debate concerning whether the use of the Transition Question should be regarded as a construct approach rather than a criterion approach due to its high face

validity, a criterion approach has been adopted in this study (Norman *et al.,* 1997). Data were collected using the BQ, RMDQ and VAS at baseline, one week, and six weeks from new patients or former patients if presenting with a new episode of acute low back pain symptoms. The Transition Question acted as the "anchor" against which responsiveness is evaluated. The Area Under the Curve (AUC) calculation derived from Receiver Operating Characteristic (ROC) curves is considered to measure the ability of an instrument to discriminate between patients who are considered to have improved or deteriorated according to the TQ.

5.3.5 Transition Question (TQ): Measuring global change in patients

Clinicians frequently face the challenge of recommending an intervention for symptomatic relief and management of specific conditions (Guyatt *et al*., 2002b). While involving the patient in their care is increasing, clinicians do not tend to rely routinely on measures of health status to judge a patient's health status (Leplège and Hunt, 1997). In this instance, the question "How are you feeling?" is more commonly used and is an implicit part of clinical practice (Wright, 2000). When trying to assess patients' other needs for information Leplège and Hunt argued that "attention might well be shifted to means of assessing (patients' viewpoint) by way of methods capable of reflecting individuals' concerns when they become ill". Feinstein referred to this type of question as a "transition question" where it can be used to rate their concerns globally using a scale of options (Feinstein, 1987; Wright, 2000). Transition questions (TQ) can be presented in a range of formats from five points on the scale to fifteen points. It had been proposed that the greater number of options increases the ability of the questions to detect nuanced change in patients. This belief, however, has been challenged by Lauridsen *et al*. who identified in their study of back pain patients that only 7% more patients were classified as improved when a 15-point scale was used compared to a 7-point TQ scale. The lack of standardisation of TQ scales was discussed and recommendations were made for the use of standardised transition questions based on this work (Lauridsen *et al*., 2007).

When measuring the outcome of care by collection of patient reported outcomes, it is important to reflect upon what those scores actually mean from a clinical

perspective, and not just from statistical evaluation. Hägg refers to the idea that a statistically significant score change does not necessarily mean that the change is clinically important (Hägg *et al*., 2003). In studies where the sample size is large enough, any minor score change could be statistically significant (Fairbank *et al*., 1980; Deyo *et al.,* 1991; Lydick and Epstein, 1993; Bombardier *et al.,* 1994; Shekelle *et al*., 1994; Deyo and Patrick, 1995; Beaton, 2000; Lydick, 2000).

The importance of the Transition Question (TQ) in measuring change

Several different approaches to the classification of meaningful change from within a clinical setting have been suggested; such approaches tend to be based upon study designs and the particular construct being measured (Husted *et al.,* 2000; Beaton *et al*., 2001; Lassere *et al*., 2001; Norman *et al.,* 2001; Wells *et al.,* 2001; Terwee *et al.,* 2003). One of the more commonly used approaches is to distinguish between distribution-based and anchor-based approaches. Distribution-based methods depend on the distribution of data, and in anchor-based methods an external "anchor" is used, *e.g.* a transition question, against which change is compared. These methods help to identify change scores which represent clinically important change from those that do not.

Anchor-based methods

The concept of Minimal Clinically Important Difference (MCID) has been introduced in longitudinal designs. This attempts to define what is the smallest meaningful change score (Jaeschke *et al*., 1989; Beaton *et al.,* 2001; Guyatt *et al.,* 2002). The term minimally important difference (MID) has appeared more commonly in the literature since 2003 to present (Walters and Brazier, 2003).

Distribution-based methods

The most common distribution-based methods are the effect size, the Reliable Change Index, and simple change scores from the outcome measures themselves (Hurst and Bolton, 2004). Effect size is a widely-used measure to assess the magnitude of treatment-related changes over time, and can be applied to individual patient data, and group data (Wyrwich and Wolinsky, 2006). It is a statistical method where the mean differences between pre-treatment and post-

treatment scores can be standardised allowing the quantification of a treatment's effect in units of standard deviation. Threshold values for effect sizes have been proposed by Cohen for group mean changes, and Testa for individual changes (Cohen, 1977; Testa, 1987).

Anchor-based and distribution-based methods have both advantages and limitations, and many authors propose using both approaches (Wright and Young, 1997; Guyatt, *et al*., 2002; Crosby *et al*., 2003). All measures of health assessment have strengths and limitations since they are used widely in mixed settings and populations. The TQ is not different and consequently there are contrasting views on whether the TQ should be used as a gold standard since a choice of cut off value can be difficult for the patient to decide upon, and the reliability of one item can be poor.

The validity of using the transition question is uncertain since patients may have difficulty in recalling their health prior to treatment, and their current health status may impact more heavily on their evaluation (Mancuso and Charlson, 1995; Guyatt *et al.,* 2002; Grøvle *et al*., 2014). Grøvle *et al*. identified that the magnitude of MIC increased when recall increased from 3-24 months. Baker notes that patients' understanding of their condition may have an effect on their recovery or symptom deterioration, and this impacts upon expectation (Baker, 1998). Patients' expectations in turn have been identified as having an effect on their perceptions of treatment benefit (Ross, 1989). The use of a retrospective measure of change is also regarded as problematic by some researchers since some patients may have forgotten aspects of their illness or symptoms. In other instances, patients may choose to report overall change in a particular way depending on whether they are satisfied or dissatisfied with their care (Baker, 1998).

In contrast, others view the TQ to have key advantages. One significant advantage is the reduced patient burden since one question is asked, and in some studies on one occasion only. The use of one question lends itself to simplified data analysis, and avoids the need to develop or administer responsive health questionnaires for individual conditions (Baker, 1998; Haspeslagh *et al*., 2006; Harris *et al*., 2016).

The TQ also offers an independent standard that is easily interpretable (Guyatt *et al*., 2002).

Notwithstanding these issues the transition question remains a useful tool, and external criterion for anchor-based methods of evaluating change. Consequently, a transition question has been included as a follow-up question in the PROM app.

5.3.6 Interpreting change

In a clinical setting it is important to examine whether a treatment effect or change is meaningful, and whether it should be advocated in either a particular patient population and/or setting. To address this issue, the concept of the Minimal Clinically Important Difference (MCID) has been introduced into the literature. This is now more commonly termed the Minimal Important Difference (MID). Jaeschke *et al*. originally defined MCID as "the smallest difference in score in the domain of interest which patients perceive as beneficial and would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (Jaescke *et al.,* 1989; Hägg *et al*., 2003). Stratford *et al.* defined the MCID as "the smallest change that is important to patients." (Jaeschke *et al.,* 1989; Stratford *et al*., 1998). At this point it is important to clarify that MCID or MID will refer to changes of importance at population level, and Minimal Clinically Important Change (MCIC) or Minimal Important Change (MIC) as it is now known; I will use this definition in my thesis following the suggestion of Terwee et al., 2003 referring to changes of importance at the level of the individual patient (Terwee *et al*., 2003; Van der Roer *et al*., 2006). To assess a patient's perspective on the interpretation of change scores, an external criterion can be used to which the score change can be compared.

Calculating MIC and MID

Different approaches to calculating the MIC and MID have been described in the literature (Jaeschke *et al.,* 1989; Lydick and Epstein, 1993; Crosby *et al*., 2003; Terwee *et al.*, 2009; de Vet *et al*., 2010; Terwee *et al*., 2010; Froud and Abel, 2014 The two most frequently used approaches to evaluating the MIC are anchor based and distribution-based approaches. Anchor based approaches use an external

criterion. One anchor-based approach is the mean change in score on the instrument in the sub category where patients are minimally importantly changed according to the anchor.

In contrast, the major disadvantage of all distribution-based methods is that they do not provide a good indication of the importance of the observed change. Crosby *et al.,* 2003 suggest that a compromise between the two approaches could be used, and de Vet *et al.,* 2007 designed a method which integrates both approaches and is called the visual anchor-based MIC distribution. More recent work by Froud and Abel focussed upon comparing methods for identifying MIC thresholds, and proposed a new approach (Froud and Abel, 2014). In summary, Terwee *et al*. identified the following basic approaches.

Distribution-based approaches:

- Using standard deviation (Norman *et al*., 2003);
- Modified Reliable Change Index (Terwee *et al*., 2010);
- Standard Error of Measurement (SEM) (Wyrwich *et al*., 1999a; Wyrwich *et al*., 1999b).

Anchor based approaches:

- Mean change method (Jaeschke *et al.,* 1989);
- MIC values based on ROC curve data (Deyo and Centor, 1986)

If we look at the different approaches in more detail, the distribution approaches will be described first.

- Norman *et al.* suggested using the formula:

  MIC = 0.5 x standard deviation of the baseline score

  This approach was selected as the value for 0.5 multiplied by the standard deviation of the baseline score is equal to an effect size of 0.5 (Norman et al., 2003; Terwee *et al.,* 2010).

- Wyrwich *et al*., 1999 suggested an approach using the SEM. It has been suggested that this is a more suitable approach since the SEM is expressed in the original metric of an instrument and is independent of the sample used in evaluation. Terwee *et al*., 2010 calculated the SEM as the square root of the error variance, and subsequently arrived at:

  MIC = 1 x SEM

- Modified Reliable Change Index (RCI). The modified version of the RCI focusses on the change in one patient. RCI is closely related to the smallest detectable change (SDC) *i.e.* 1.96 x √2 x SEM, and the modification allows better distinction between individual and group level analysis. It is calculated by:

$$\frac{\text{(The change score) of one patient} + \text{(mean change in) the whole group of patients} \times (1 - \text{reliability of the change score})}{\sqrt{\text{reliability of the change score}}\ \sqrt{2}\ \text{SEM}^2}$$

Anchor based approaches:

- Mean change method (Jaeschke *et al.,* 1989)
  Based on this approach, the MIC can be calculated as:
  Mean change score in the subgroup of patients reported as a "little better" on the anchor $\mp$ $\frac{\text{(Standard Deviation of change)}}{\sqrt{n}}$

- MIC values based on ROC curve data (Deyo and Centor, 1986).

  In this approach, the MIC is defined as the smallest amount of mis-classification. Using the anchor *e.g.* a global rating of change, patients can be divided into those reporting improvement or no change. The distribution of the change scores are plotted. Sensitivity, and 1-specificity are then plotted to obtain a ROC curve. de Vet *et al.* recommended an integrated approach which encapsulated distribution and anchor-based methods and is known as the visual anchor-based MIC distribution (de Vet

et al., 2007). In this approach the change scores of three groups are plotted. The three groups concerned are patients who have importantly improved, not importantly changed, and importantly deteriorated. Once values are plotted, a cut off value leading to minimal misclassification is sought. The optimal cut-off point which minimises the sum of 1-specificity and 1-sensitivity is regarded as the MIC value: this point is on the upper left corner of the curve.

- Farrar *et al.* choose the point closest to the intersection of a -45° tangent line. Mathematically this is equivalent to the point where sensitivity and specificity are closest together (Farrar *et al.,* 2001).
- Froud and Abel suggested a new method based on the addition of the sum of the squares of the 1-sensitivity and 1-specificity. This is in accordance with Pythagoras' theorem (Froud and Abel, 2014).

Froud and Abel argue that choice of MIC estimator is important due to its potential on interpretation of trial findings. In previous research, values for MIC have been calculated for the BQ, the VAS, and the RMDQ. These will be used for comparison to the MIC values calculated in this study when using these outcome measures in an electronic format, and in an osteopathic practice setting (Ostelo *et al*., 2008; Newell and Bolton, 2010).

MIC and SDC

An additional measure of interpretability is the Smallest Detectable Change (SDC) also termed the Minimal Detectable Change (MDC), and this will be calculated for all PROMs included in the app. The values for SDC will be compared with MIC. It should be stressed that the MIC and the SDC are not the same. SDC is based on measurement error, and is the change beyond measurement error. It is closely related to measurement error and the reliability of instruments.

SDC = $\pm$ 1.96 X SD of pre-test – post-test scores
= $\pm \sqrt{2}$ X SEM (Beckerman *et al*., 2001)

SDC is similar to the Reliable Change Index = $\frac{\text{pre-test – post test scores}}{\text{SD of the difference}}$

SD of the difference = $\sqrt{2}$ X SEM

SDC = 1.96 x $\sqrt{2}$ x SEM (Terwee et al., 2009)

Change scores smaller than the SDC can be attributed to measurement error (ME). The relationship of SDC with measurement error implies that when an instrument has small ME, relatively small changes can be identified as small changes. If ME is large, changes must be large to be sure changes are not due to ME. When calculating the SDC, the SEM used should be based on test-retest parameters and not on Cronbach α as this does not assess change in scores at different time points. Reasons for variations in patients completing instruments include biological variation in patients, their mood while completing the questionnaire, measurement variation in the observer who may apply different criteria to the questions or completing a measurement instruments on different days. There could also be systematic error occurring, for example, if all participants changed slightly by suffering the effects of a cold, or a local change in the weather affecting all participants' responses. It is for this reason that test-retest parameters are used for measuring measurement error.

Measurement error reduces with repeated measures, thereby smaller changes beyond measurement error can be detected *i.e.* SDC becomes smaller. ME is reduced when measuring in groups of patients, so SDC is reduced by a factor of $\sqrt{n}$ when a group of n patients are studied.

5.3.7 Measuring patient satisfaction and experience

I described in Chapter 1 how the role of the patient has changed within healthcare. Due to its significance within quality standards and clinical governance, patient satisfaction and experience are measures of change included in the app. It could be argued that patient satisfaction is being used as a form of proxy for quality of care, but in effect patient satisfaction is a complex multidimensional construct (de Almeida *et al.*, 2015). Although extensive and complex questionnaires are available that measure patient satisfaction and experience alone, their appropriateness depends on the purpose of measuring satisfaction and experience and the settings within which they will be used (Hekkert *et al.*, 2009; Hibbard *et al.*, 2010; Fenton *et al.*, 2012). Over the past 10 years, there has been a gradual shift in

emphasis from measuring patient satisfaction to that of measuring patients' experience of their care. This has been driven by the recognition that while patient satisfaction measures can be fairly insensitive in detecting shortcomings in care, they are affected also by patients' expectations which could be low in some populations (Black and Jenkinson, 2009).

Patients offer complementary perspectives to care to that of their clinicians, and provide unique insight into the experience of that care, and its outcomes (Black and Jenkinson, 2009). The value of using patient satisfaction data alone is disputed by many commentators emphasising the views of Black and Jenkinson, 2009 concerning its sensitivity to identify shortcomings in care (Black and Jenkinson, 2009; Fenton *et al*., 2012; Greaves *et al*., 2012). Some of the dissonance associated with this measure is that patient satisfaction is essentially a judgement about whether expectations have been met, and this is influenced by varying standards, the patient's general wellbeing, their waiting time for treatment or consultation, and perhaps their previous experience of an institution or individual in addition to the low expectations cited by Black and Jenkinson (Devkaran, 2014). Evidence is limited concerning how patient satisfaction data is used to support quality improvement, and increased levels of satisfaction are not necessarily linked with quality improvement (Carr-Hill, 1992; Sitzia and Wood, 1997; Hudak and Wright, 2003; Crow *et al*., 2002). One explanation for this apparent dissonance is the view that satisfaction is used as a process measure and at other times as an outcome measure which is perhaps a reflection of how satisfaction measurement developed originally in response to marketing and efficiency evaluations during the introduction of free-markets into healthcare provision (Salisbury *et al*., 2010).

Patient experience extends beyond satisfaction with care, and can also embrace occasions when the outcome is negative but the experience has been good, and vice versa. It is becoming increasingly apparent that patients judge healthcare delivery on compassionate and engaged care delivered by clinicians, and are not solely concerned with outcome of treatment (Devkaran, 2014). This represents an interesting perspective when considering the work by Pincus *et al*. who identified high levels of satisfaction despite lower outcomes in their osteopathic care

provision (Pincus *et al*., 1992). Although satisfaction is described as a difficult concept to describe, its multifaceted nature suggests that it should not be dismissed in favour of other measures. It is perhaps more pertinent to ensure that the reason for identifying levels of satisfaction is clear, and this may provide clarity to the most appropriate questions used in its evaluation.

A range of initiatives have been introduced to measure patients' experience of care and the care setting. This has been manifest in the NHS innovation known as the Friends and Family Test (FFT) which has been introduced into acute care and mental health services throughout 2014 (https://www.england.nhs.uk/ourwork/pe/fft/). Meeting patients' expectations to ensure high levels of satisfaction and experience is challenging in a context where expectations of treatment and healthcare professionals are high (Avis *et al*., 2008; Lateef, 2011). However, satisfaction, experience, and expectation are closely related to each other.

A substantial review of experience, expectations and satisfaction in both commercial and healthcare contexts (Thompson and Sunol 1995) described how expectations are defined variously: in psychological terms as cognitive beliefs, in sociological terms as predictors of social interaction, and in biomedical research as the mechanism of the placebo effect, and as a component of satisfaction, with satisfaction being a goal of healthcare management. The authors also identify why expectations within healthcare are rather different to expectations in other arenas of service provision such as hotel services or retail.

In healthcare, conscious expectations are often unformed, especially in the first-time patient, because of the esoteric or technical nature of the treatment. Instead, expectations are 'epi-phenomenal', emerging as part of a dynamic interaction between provider and user. Expectations also have a strong affective (as well as cognitive) component in healthcare, because of the emotional and literally 'extra-ordinary nature of the interaction; the long time-course of the interaction if the disease is chronic; the emotion associated with suffering and pain; and the fact that a patient's needs are rarely trivial (in contrast to the desires of a consumer). The

primary focus of the interaction is on reducing those needs (in contrast to increasing demand as in marketing). Thompson and Sunol point out the intimacy of healthcare interactions, when there is often sanction for the patient to be physically and/or psychologically 'laid bare'. Finally, the emotional charge of a successful outcome, of triumph over adversity, can remove from memory some of the dissatisfactions, poor experiences, and disappointments encountered along the way, in other words can cause *post-hoc* re-evaluation of negative perceptions of the service (Thompson and Sunol, 1995).

A dynamic model of how patients perceive non-pharmacological treatments such as manual therapy (Yardley *et al*., 2001) confirmed the dynamic 'epi-phenomenal' nature of expectation in healthcare: they found an interplay between the patients' values and beliefs about illness and treatment, their experience of the therapist and the therapy, and their subjective experience of change in symptoms. Initial beliefs and expectations, including beliefs about the appropriate type of treatment, can be modified over time. Symptom improvement can increase trust and congruence between the illness-treatment models of patient and therapist. A second qualitative study found that patients who were off sick due to musculoskeletal disorders often had no prior expectations of the treatment on offer (which included a cognitive-behavioural component), yet half declined treatment because they felt that it was 'too psychological' and did not match their beliefs about what was appropriate for their condition. A study in physiotherapy showed how the therapist can modify over time their patients beliefs, such as acceptance of their back pain problem and desire to be involved in management (May, 2003).

Unmet expectations and poor experiences of care can lead to dissatisfaction, although several studies conducted in the USA in primary care have focussed on a limited, 'biomedical' set of expectations (Jackson and Kroenke 2001; Jackson *et al.*, 2001; Bell and Kravitz 2002) such as a diagnosis, a diagnostic test, a prescription, diagnostic information and prognostic information. Satisfaction and trust can remain high even if these biomedical expectations are unmet (Keitz *et al*., 2007). A study in UK primary care (Williams *et al.,* 1995) suggested that patients wanted an

explanation of the problem more than tests or diagnosis. This desire for understanding of the complaint emerges in many studies (Cooke *et al.*, 2006). In the consultation, patients seem to focus on biomedical issues; other concerns such as worries about the complaint, and social issues are often unvoiced in the consultation, which can lead to poor outcomes and lack of compliance (Barry *et al.*, 2000).

Some inconsistent results emphasise that the relationship between expectations, experiences, and outcomes is complex (Sherman *et al.,* 2010); one possible factor being the failure of standard outcome measures to capture outcomes that patients value such as hope, relaxation, ability to cope and general well-being (Hsu *et al.*, 2010). Hope of recovery is not well researched, and is best understood as expectation of recovery (Wiles *et al.*, 2009) and as a potential positive force that can be encouraged by health professionals (provided they do not encourage false hope). The inadequacy of standard outcome measures was further demonstrated in a qualitative analysis of patients' free text responses in the BEAM trial (Underwood *et al.,* 2006) which suggested that patients perceived much greater benefits from the manipulative arm than from usual care, a difference which was not detected by the quantitative analysis of standard outcome measures.

Patients' expectations and experiences of treatment can be influenced, or usefully increased, by the clinician: in patients with minor ailments; the physician providing a favourable prognosis was linked to better outcome for the patient, but only if an explanation of the complaint was also given (Fassaert *et al.*, 2008). Surgeons' overly optimistic expectations about improvements in quality of life due to surgery for back pain did not show a significant '*curabo*' effect on outcomes generally, but did improve psychological dimensions in a sub-group of these patients (Graz *et al*., 2005). Conversely, patient expectations can influence the therapeutic actions of the clinician. This has been much researched in relation to prescribing of antibiotics (Macfarlane *et al.*, 1997). While patient need is the strongest factor for determining prescribing of tests, referral or medication by GPs, perceived patient pressure is also a strong independent predictor (Little *et al*., 2004).

The context of the healthcare service also influences expectations and experience. Primary care represents the front line of care and patients' expectations of primary care today (in the NHS) were well summarised by Coulter (Coulter, 2005). The most important expectation was 'humaneness' or interpersonal care, followed by competence, involvement in decisions, and time for care. GPs in the UK tend to paternalism, scoring high on communication, continuity and affordability, but low on information and choice. Patients also value fast access to services, and clear information for self-management.

Patients with unexplained physical symptoms value the time and listening they receive from Complementary and Alternative Medicine (CAM) practitioners, as well as an interactive and holistic approach (Rugg *et al*., 2011). Beliefs tend to vary in different illness groups (Bishop *et al.*, 2007). The beliefs and expectations of users of different types of CAM therapy when compared to patients in general practice showed that, of the various CAM users, osteopathic patients were most similar to the GPs' patients (Furnham *et al.,* 1995). Users of a multi-therapy CAM service provided by the NHS (Richardson, 2004) expected symptom relief, improved quality of life, information, advice, a holistic approach, and wide availability of such therapies on the NHS.

When seeking private care, patients make choices about the type of care they feel appropriate and choose their practitioner: they become health care consumers. The way they make these choices is not straightforward, especially outside of conventional healthcare: personal recommendation from trusted others, the practitioner's trustworthiness and technical competence are important (Bishop *et al.,* 2011a). As consumers they expect value for money together with "added extras" in the environment, good access to care, and individualised holistic care (Bishop *et al*., 2011b). For patients who are naïve to the therapy, the decision to return after the first visit depends on both relief of symptoms and the relationship they have established with the therapist (Grimmer *et al.,* 1999). Patients with long term conditions appear to try to find a therapist who will legitimate their condition, and tailor their consultation strategies through their illness pathway (Parsons *et al*., 2011). Shared decision making/active participation in treatment

had a positive effect on outcomes. Patients seeking private osteopathic care were found to manage their search for health, they value the interpersonal relationship and hope of symptom improvement; and they act as consumers, benchmarking the quality of the service and professional expertise against NHS services (Strutt *et al.*, 2008).

Although increasing numbers of questionnaires are becoming available that measure patient reported experience, *i.e.* Patient Reported Experience Measures (PREMs), and patient satisfaction, the inclusion of a complex PREM and patient satisfaction questionnaire in addition to the PROMs and other demographic questions seemed inappropriate when considering the comments from the patients that completion of the app should be fairly quick.

The multi-faceted issues present in a patient's consultation and treatment could not be captured in a Patient Reported Outcome Measure alone. It is precisely for this reason that measures for patient satisfaction and experience are a vital part of measuring quality in healthcare. Single questions to measure patient satisfaction, and patient experience used extensively by the Picker Institute were selected for inclusion in the app (http://www.pickereurope.org/).

5.3.8 Data security and patient confidentiality

Once the content of the app had been agreed, and the functionality tested, the important area of data security needed to be assessed. While the development of apps to collect patient data provides new opportunities and benefits, they may also expose patients to potential new risks (Steinhubel *et al.*, 2013; Kotz, 2011; Njie, 2015). To increase the use of apps in healthcare settings, patients must feel confident in the skills of the developers that their privacy is protected, and their data is used in an ethical manner pre-defined prior to use (Cohn, 2006; Smith *et al.*, 2011). In the UK, the Data Protection Act (1998) enshrines eight principles which limit the appropriate and proportionate collection of personal information. IT requires that the use of personal data are clearly stated in a privacy policy, establishes the rights of an individual to control the use of their data, the right of an individual to amend their data should they wish, and mandate safeguards

against situations where their data might be accessed and their privacy potentially be compromised (HM Government, 1998). While the manner in which data are collected and used are of concern in the use of apps, there are also issues relating to data transfer and unauthorised access. The use of encryption for data transfer should be regarded as standard in the UK, and safeguards should be in place for data storage to prevent malicious access of data by external individuals or groups *e.g.* hacking (Dehling *et al.,* 2015). There are, however, international differences concerning the use of encryption. The export of encryption products from the United States is regulated by a variety of governmental agencies. The primary regulator of encryption exports is the Commerce Department's Bureau of Industry and Security (BIS), which administers the Export Administration Regulations (EAR) (Saper, 2013). There are specific requirements for data relating to medical use.

Accreditation programmes have been introduced to try and address some of the issues described and provide confidence for patients when using health apps. The NHS Health Apps Library was created in 2013 to offer a list of suitable apps which had undergone appraisal processes and provide greater confidence in their standards for healthcare professionals. The Health Apps Library is required by the Office of the Information Commissioner (ICO) to provide information about registration with the ICO, to declare any data transmissions by an app, and possess a privacy policy for patients. In a systematic assessment of the apps within the Health Apps Library, Huckvale *et al*. reviewed "compliance with recommended practice for information collection, transmission, and mobile device storage" (Huckvale *et al.,* 2015). Confidentiality arrangements, and the availability of privacy policies were reviewed also. The findings of the review caused considerable consternation since 89% of apps transmitted information to online services but no apps encrypted data when stored locally, 66% of apps sent identifying information over the internet, and only 67% had some form of privacy policy. These issues were addressed prior to the PROM app being prepared for piloting. Data security was tested by Clinvivo Ltd to assess the app's ability to resist hacking. Privacy of information and its use was assured as this had been a key issue raised by patients in the interviews when developing the app (Chapter

2). All data are encrypted end to end using industry standard Secure Sockets Layer (SSL). Data are encrypted also when sent from Clinvivo to me for analysis.

5.4 Methods: piloting the app

Once initial content and follow up periods for the app had been agreed, α- and β-testing had been completed, and data security issues had been assessed, the piloting process could begin using a larger number of volunteer osteopaths. A range of stages including recruitment, piloting of the app in a range of settings, and data collection are described in the next section.

5.4.1 Aims of the pilot process

- To identify the feasibility of using practice-based data collection in osteopaths' private practices;
- To identify practical issues experienced by patients and clinicians in practice-based data collection;
- To evaluate the test-retest reliability of the outcome measures included in the data collection app;
- To evaluate the responsiveness of the outcome measures included in the data collection app;
- To identify lessons from patients and clinicians about the process to inform wider scale implementation of practice-based data collection.

5.4.2 Study design

This pilot study involves three distinct phases. These are shown in Figure 5.7.

5.4.3 Recruitment of osteopaths

Osteopaths were recruited using convenience sampling which has been described also as haphazard sampling (Etikan *et al.,* 2016). Members of the target population meet desired criteria which may be based upon accessibility, geographic proximity or ease of travel, and availability at particular times of the day (Dörnyei, 2007; Etikan *et al.,* 2016). Once osteopaths had agreed to participate in the pilot project, they were asked more questions about their practice.

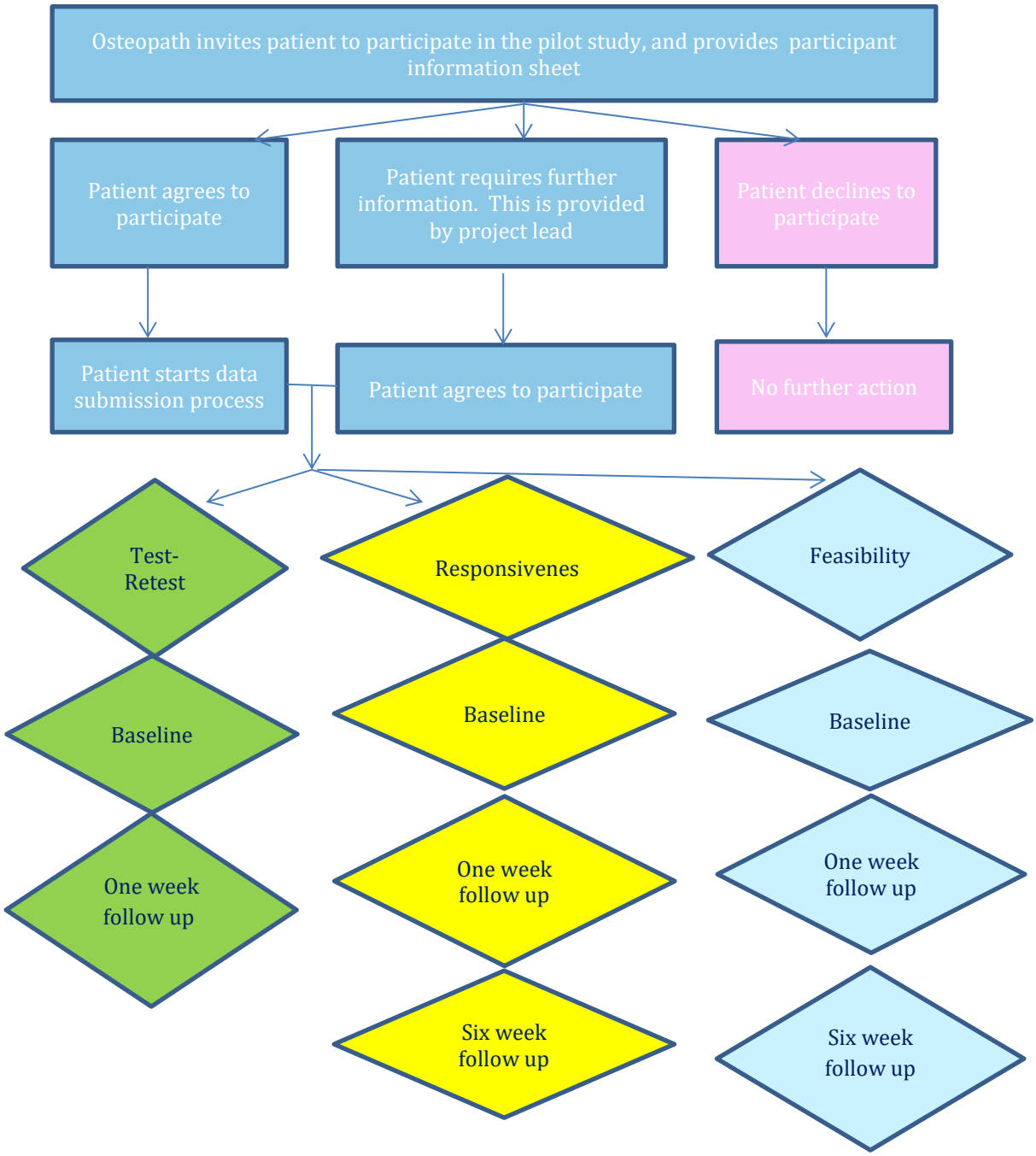


Figure 5.7 Pilot study process

This was to establish whether they saw more patients with acute or chronic symptoms and identify the arm of the pilot study for which their patients would be most suitable.

5.4.4 Recruitment of patients including sample size

As shown in Figure 5.6, there were three separate arms to this pilot study and patients were included in the appropriate arm based on the duration of their

symptoms. For example, patients with symptoms of low back pain of duration 0-6 weeks were included in the responsiveness arm of the trial, patients with chronic, stable symptoms of low back pain of duration 13 weeks or more were included in the test- retest arm of the study, and patients with chronic changeable or subacute symptoms were included in the feasibility arm of the study.

Patients suitable for each respective arm of the study were invited on a consecutive basis by the clinician to participate in the PROM pilot. They were provided with the relevant participant information sheet, a code, and information concerning how to access the PROM app.

5.4.5 Settings and locations for data collection

Data collection took place in private practices throughout England. No osteopaths in Scotland, Wales, or Northern Ireland volunteered for the pilot. Data collection could not take place on NHS sites or where NHS patients were involved due to the constraints of the ethics permission for the study.

5.4.6 Ethics and governance

This study involved patient and clinician participants. It was clear that this was a research study as opposed to a different methodological approach. Ethical issues were considered including the rights and wellbeing of all of the participants, and the study was conducted according to Good Clinical Practice (Research Governance Framework, 2005; HRA, 2016).

A participant information leaflet was provided to interested osteopaths concerning their recruitment of patients into one specific strand of the pilot project. A consent form was included with the leaflet, and this was returned prior to patient recruitment beginning. Separate participant information sheets were created for each strand of the pilot project for either the test-retest, responsiveness or feasibility arms of the study. They were tested for readability and accessibility across a range of reading ages achieving a Flesch Kincaid reading ease score of 76.9.

These leaflets were clearly marked designating their purpose and gave information to patients about what the study would involve, the purpose of the data collection, and reassurances concerning confidentiality and privacy in respect of the use of their data (Appendix 5.1, 5.2, and 5.3). No separate consent form was created since completion of the questionnaire within the app was deemed to be evidence of consent (Ploug and Holm, 2015; HRA, 2016). Patients were also offered £5 in the form of a "Love to Shop" voucher on completion of the questionnaires.

Ethics approval for the pilot study was obtained from the research ethics committee at Queen Mary University of London (QMERC2014/18). Clinicians working in private practice only were included in the study; the ethics permission did not extend to NHS clinicians while working on NHS sites.

5.4.7 Outcome measures used

A range of data was collected using the PROM app. This included demographic data (age, sex, ethnic background, and work status), service data concerning access to appointments, the duration of symptoms, the primary reason for seeking treatment, and the site of all symptoms. Three outcome measures were included *i.e.* the Roland Morris Disability Questionnaire (RMDQ), the Bournemouth Questionnaire (BQ), and a visual analogue scale (VAS). At follow up, one week and six weeks after the initial questionnaire had been completed, patients were asked to complete the BQ, RMDQ, and VAS. In addition, they were asked to complete questions concerning patient experience, patient satisfaction, and a transition question.

Changes to trial outcomes after the trial commenced with reasons

One minor change to the questionnaire was made during the pilot process. On examination of some of the early data I observed that few patients were completing the RMDQ. While patients are required to tick the questions in the RMDQ which relate to their symptoms, it was unclear whether patients were not completing the RMDQ because they did not have symptoms, or because they had bypassed the questions. To clarify this situation, an extra mandatory box was

included by the Clinvivo team to denote that the patient had "none of the above symptoms" in relation to the list of RMDQ questions.

5.4.8 Statistical methods used to compare groups for primary and secondary endpoints

Data were transferred to me from Clinvivo using an encrypted process involving Gpg4win encryption software and Kleopatra (Intevation GmbH, Osnabrück, Germany) which is a certificate manager and a universal crypto graphical user interface. These data were transferred to an Excel spreadsheet at my request to allow data analysis in Excel, Statistical Package for the Social Sciences (SPSS; IBM, Washington) version 22, and small STATA14 (StataCorp LP, Texas). A variety of statistical tests were identified prior to data collection: they included descriptive statistics for the demographic and service data.

Participants provided sociodemographic information at baseline, and change data at baseline, one week, and six weeks. These data were evaluated:

Baseline sociodemographic data

Sociodemographic data will be presented in tabular and graphical format displaying baseline characteristics, with frequency, and confidence intervals for each characteristic. The duration of patients' current symptoms, the main areas affected by symptoms, the number of symptom areas reported, and their evaluation of general health status will be presented also as frequency distributions.

Service data

Patients' waiting time until the first appointment offered, evaluation of satisfaction and experience will be presented as a frequency distribution, The relationship of satisfaction and experience to demographic characteristics and outcome will be evaluated also. The statistic test will be dependent upon the comparison of categorical and continuous data in the combinations relevant to the characteristics and outcomes being examined.

Pain intensity

Pain intensity has been measured using a visual analogue scale. Data concerning change in VAS scores have been evaluated in a range of ways including:

- The raw scores at pre- and 6 week post-treatment stages (Little and MacDonald, 1994; Farrar *et al.,* 2001; Hurst and Bolton, 2004; Gurden *et al*., 2012);
- Percentage change score (Little and MacDonald, 1994; Farrar *et al*., 2001; Hurst and Bolton, 2004; Gurden *et al*., 2012);
- The number of individual improvements and deteriorations;
- Area under the curve values from plotting ROC curves (Deyo and Centor, 1986).

Pain and disability

Pain and disability was evaluated, with other domains of interest, using the Bournemouth Questionnaire (BQ) and the Roland Morris Disability Questionnaire (RMDQ). Regression analysis examined the relationship between sociodemographic characteristics and changes for the outcome measures. Change scores for the RMDQ and BQ have been evaluated including:

- The raw scores at pre- and 6 week post-treatment stages (Little and MacDonald, 1994; Farrar *et al.,* 2001; Hurst and Bolton, 2004; Gurden *et al*., 2012);
- Percentage change score (Little and MacDonald, 1994; Farrar *et al*., 2001; Hurst and Bolton, 2004; Gurden *et al*., 2012);
- The number of individual improvements and deteriorations;
- Area under the curve values from plotting ROC curves (Deyo and Centor, 1986).

Transition Question

Although there is discussion in the literature concerning the merits of using global health transition questions compared to domain-specific transition questions, global change has been measured. Data from the Transition Question were examined and acted as an anchor to evaluate measures of responsiveness (Ward *et al*., 2015). The Transition Question was used at one week and six weeks and

output was arranged into dichotomous change *i.e.* change or no change. Participants' scores of "change" were defined as including "completely recovered" and "much improved". All other scores were defined as "no change". Since the data from the PROMs were continuous, values for each sets of scores were plotted to produce a Receiver Operating Characteristics (ROC) curve and the Area Under the Curve (AUC) was evaluated for each PROM. At analysis, patients who had recorded "no change" on the transition question were included in analysis for calculating the smallest detectable change (SDC). The SDC is based on measurement error and is calculated using the formula:

SDC = 1.96 x $\sqrt{2}$ x SEM where SEM is the standard error of measurement (Terwee et al., 2009).

Test-retest reliability of the outcomes measured

Reliability has defined already by Mokkink *et al.* (Mokkink *et al.,* 2010a). However, other common terms are used which include: reproducibility, repeatability, precision, variability, consistency, stability, and agreement. Internal consistency, reliability and measurement error are all considered to be aspects of reliability (COSMIN definitions). Reliability was evaluated for the BQ and the RMDQ using an intra-class correlation coefficient (ICC) calculation, and evaluating Cronbach α for each PROM.

Reliability and measurement error

The standard error of measurement (SEM) is recognised as a parameter of measurement error. SEM is a measure of how far apart outcomes of repeated measures are: it is the SD around a single measurement. SEM is calculated based on values for the ICC. SEM will be calculated using:

$$SEM = SD_{t1/t2} \times \sqrt{(1\text{-}ICC)}$$

Responsiveness of the outcomes measured

Assessing whether a disease or symptom state has changed in a patient to whom treatment is given is one of the most important objectives in clinical care. The aim of measurement instruments is to assess whether care is making a difference to

the patient. It is this capacity that is articulated in the concept of responsiveness, and was described by Mokkink *et al.* as "the ability of an instrument to detect change over time in the construct to be measured" (Mokkink *et al*., 2010a).

Appropriate and inappropriate measures of responsiveness

The COSMIN group has extensively studied measurement of responsiveness (de Vet *et al.,* 2003). A Receiver Operating Characteristics (ROC) curve is the recommended statistical parameter to be used when the gold standard is a dichotomous variable (*i.e.* change or no change). The AUC is considered to measure the ability of an instrument to discriminate between patients who have improved or deteriorated and those who report improvement or deterioration when using the gold standard. An AUC of 0.7 is generally regarded as being appropriate. The ROC method must include patients who show change and no change (de Vet *et al.,* 2011).

A range of measures of responsiveness are reported in the literature and some are regarded as appropriate by the COSMIN group and others as inappropriate. Inappropriate measures include:

- Effect sizes Cohen (1977);
- Paired t-test;
- Guyatt's responsiveness ratio (Guyatt *et al*. in 1987).

Interpretation of measures of change

Interpretability has been defined as "the degree to which one can assign qualitative meaning *i.e.* to an instrument's quantitative scores or change scores". Terwee *et al.* suggest that 'only a change larger than the measurement error can be considered as "real" change' (Terwee *et al*., 2009). As a consequence there is a need to determine the minimal amount in change in score that patients or clinicians consider to be important. This definition of Minimal Important Chance or MIC is proposed by the COSMIN group as "the smallest change in score in the construct to be measured which patients perceive as important".

For clinicians, MIC may be one that indicates a change in the treatment or prognosis of the patient. In studies with patients it is important to determine whether results are statistically significant and clinically relevant. In the latter MIC may be of interest. For these reasons, it is important to reflect upon the MIC in relation to the measurement error of an outcome measure. Although measurement error has been discussed within the context of reliability, it is an important consideration in evaluating MIC. If the measurement error is smaller than the MIC, it is possible to distinguish change that is clinically important from measurement error and have a large degree of certainty in that evaluation. However, in situations where measurement error is larger than the MIC value, it becomes much more challenging to try to distinguish changes as large as the MIC from what may be due to change due to measurement error alone (Terwee *et al.,* 2009). The merits and demerits of the different approaches will be expanded upon in the discussion section.

In summary, the statistical evaluation allows profiling of the osteopathic patient population participating in the three strands of this study, and allows calculation of symptom change for this population using a range of different outcomes, and evaluation of the measurement properties of the included outcome measures when used in an electronic format.

5.5 Results

The pilot of the PROM app gathered a considerable amount of data. The analysis plan was described in section 5.4.8, and the findings are described in this section of the chapter. The descriptive statistics will be described first (5.5.1), and the data pertaining to outcomes will be provided next (5.5.2). Analysis of the measurement properties of the PROMs included in the app will be considered (5.5.3), and the interpretation of those measurement properties will be discussed last (5.5.4).

5.5.1 Descriptive statistics

Participants

The participants in this part of the study can be characterised into two types: the osteopaths who volunteered to recruit patients, and the patients who agreed to submit data using the PROM app.

The final number of osteopaths who collected data was less than the number of initial volunteers. The recruitment flow is shown in Figure 5.8.

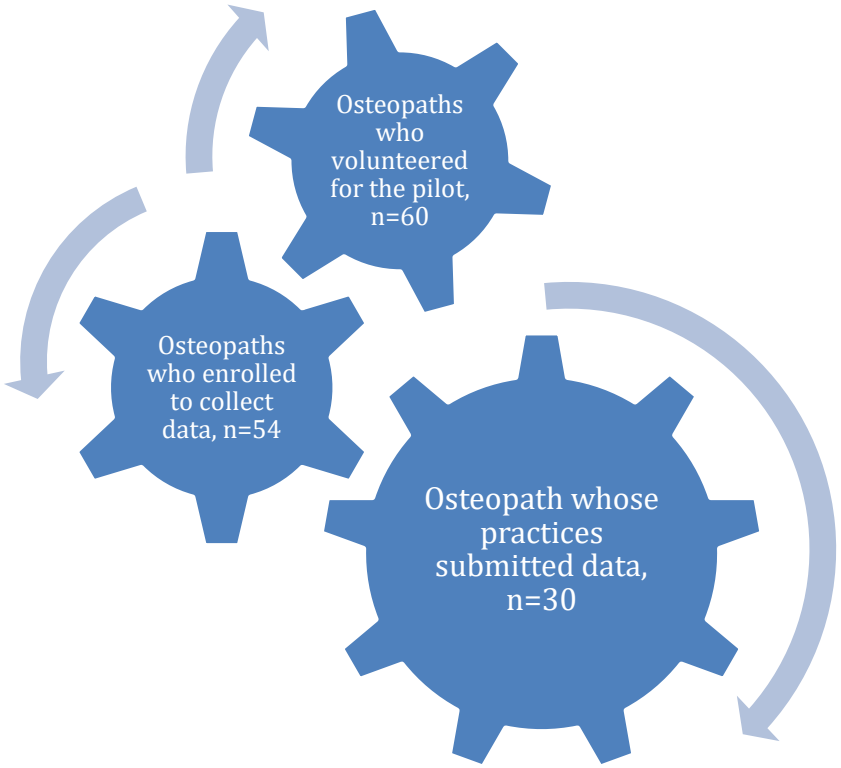


Figure 5.8. Osteopathic volunteers in the PROM pilot.

The osteopaths who participated in the pilot included 17 females and 13 males. Their mean time since qualification was 22 years, and they were distributed around the UK in Wales (n=1), Scotland (n=1), the South West (n=4), the North East (n=2), the South East (n=14), Yorkshire (n=5) and the Midlands (n=3). The distribution of participating osteopaths between each strand of the study is shown in Table 5.4.

Table 5.4. The number of participating osteopaths involved in each project strand

	Responsiveness	Feasibility	Test-retest reliability
Number of osteopaths	10	5	15

The numbers of participants included in each arm are shown in Figure 5.9.

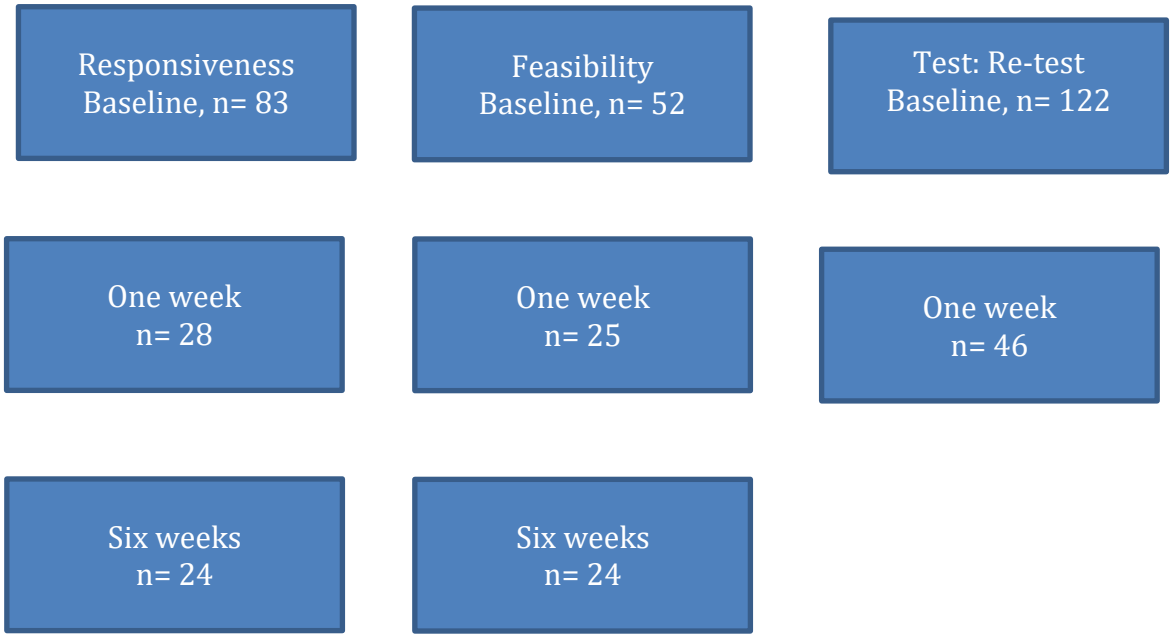


Figure 5.9. Participants in each arm of the PROM pilot

The characteristics of the patient participants in the PROM pilot are described in Table 5.5.

Losses and exclusions

Two patients were formally lost to the study. Notification came via emails to Clinvivo stating one participant would be unable to complete their follow up questionnaires due to being away unexpectedly. The second participant stated that their symptoms did not fit with the content of the questionnaire, and they did not feel they could submit relevant data. No participants asked for their data to be removed from the dataset.

In addition to the data from the demographic questions shown in Table 5.5, a range of other information was sought from patients. The findings to each of the questions are displayed graphically.

Table 5.5 Characteristics of patient participants in the PROM pilot

	Test-retest reliability	Feasibility	Responsiveness
Age			
18-29	10 (8.2%)	2 (3.8%)	15 (18.1%)
30-39	20 (16.4%)	6 (11.5%)	15 (18.1%)
40-49	26 (21.3%)	10 (19.2%)	14 (16.9%)
50-59	27 (22.1%)	13 (25%)	17 (20.5%)
60-69	25 (20.5%)	14 (26.9%)	4 (4.8%)
70-79	11 (9.0%)	6 (11.5%)	8 (9.6%)
80-89	3 (2.5%)	1 (1.9%)	0 (0%)
90 or over	0 (0%)	0 (0%)	0 (0%)
Employment status			
Employed	78 (63.9%)	35 (67.3%)	61 (73%)
Unemployed	5 (4.1%)	0 (0%)	2 (2.4%)
Retired	26 (21.3%)	12 (23.1%)	16 (19.3%)
Looking after home/family	6 (4.9%)	2 (3.8%)	3 (3.6%)
School or education	4 (3.3%	0 (0%)	0 (0%)
Long term sickness	0 (0%)	1 (1.9%)	1 (1.2%)
Other	3 (2.5%)	2 (3.8%)	0 (0%)
Ethnicity			
White (British/Irish)	99 (81.1%)	43 (82.7%)	63 (75.9%)
White (other)	12 (9.8%)	6 (11.5%)	7 (8.4%)
Asian/Asian British	6 (4.9%)	0 (0%)	1 (1.2%)
Black/African/Caribbean/Black British	1 (0.8%)	2 (3.8%)	4 (4.8%
Mixed/multiple ethnic groups	1 (0.8%)	0 (0%)	4 (4.8%)
Other ethnic groups	3 (2.5%)	1 (1.9%)	4 (4.8%)
Sex			
Male	53 (43.4%)	13 (25%)	37 (44.6%)
Female	69 (56.6%)	39 (75%)	46 (55.4%)

Age

Patients reported their age to be between 18 and 89. Histograms for age groups in each strand of the study are shown in Figures 5.10 (test-retest reliability), 5.11 (feasibility), and 5.12 (responsiveness).

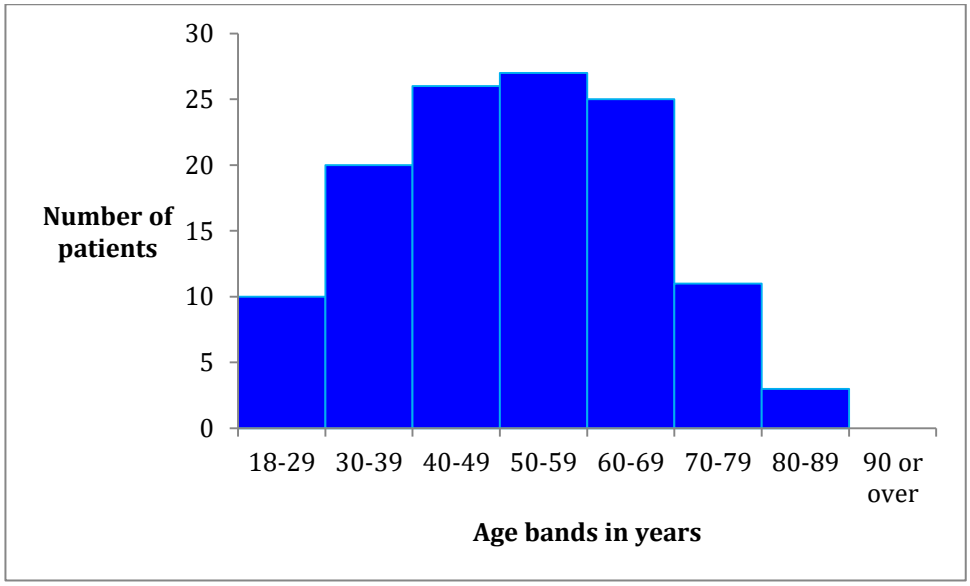


Figure 5.10. Age of participants in the test-retest strand of the PROM pilot.

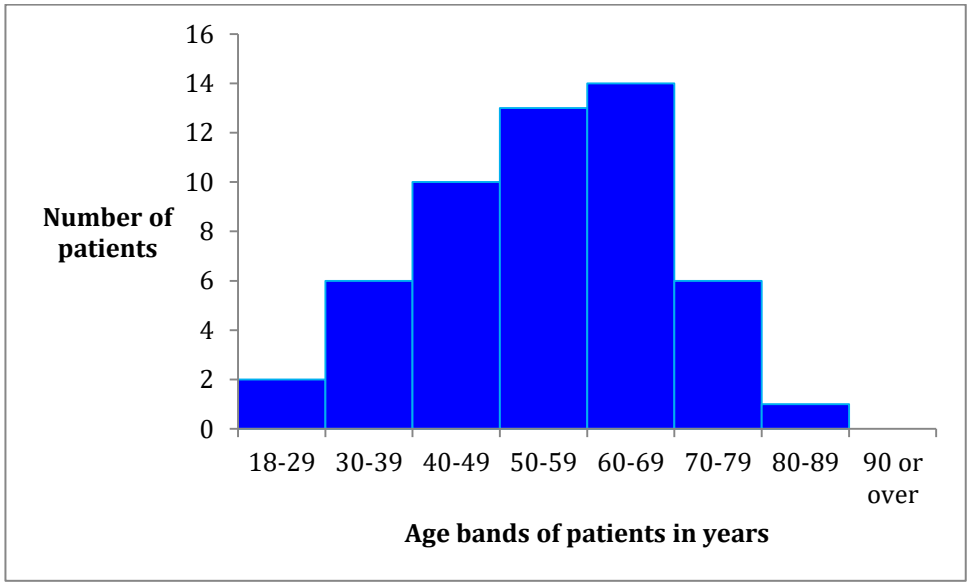


Figure 5.11. Age of participants in the feasibility strand of the PROM pilot.

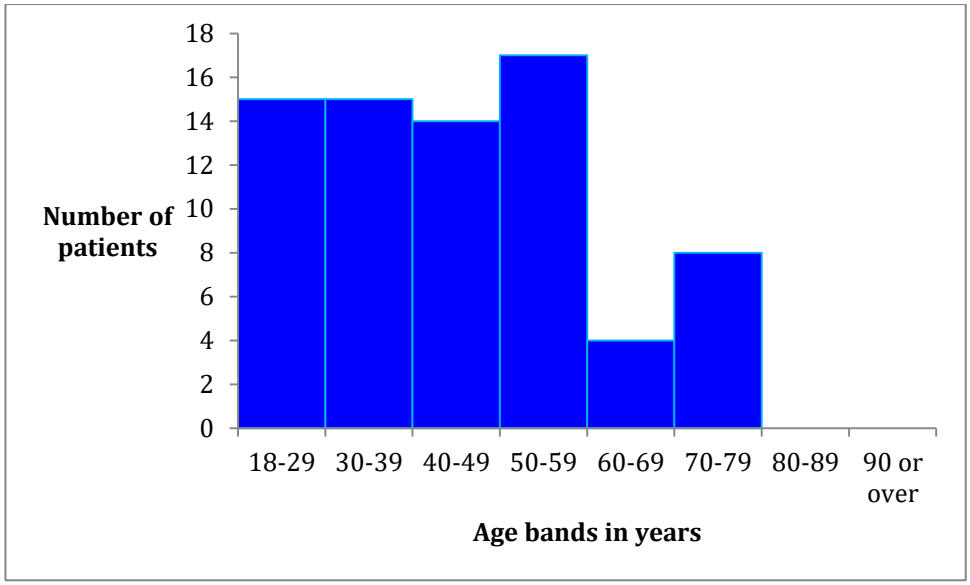


Figure 5.12. Age of participants in the responsiveness strand of the PROM pilot.

The oldest patients were represented in the test-retest strand of the pilot.

Employment status

The employment status of the participants in each strand of the pilot is shown in Figure 5.13. Employed patients accounted for the largest number of participants (67.7%), with 2.7% unemployed, 1.6% in education, 0.4% reporting long term sickness, 21% retired, 4.3% looking after a home/family, and 1.9% who described themselves as "other".

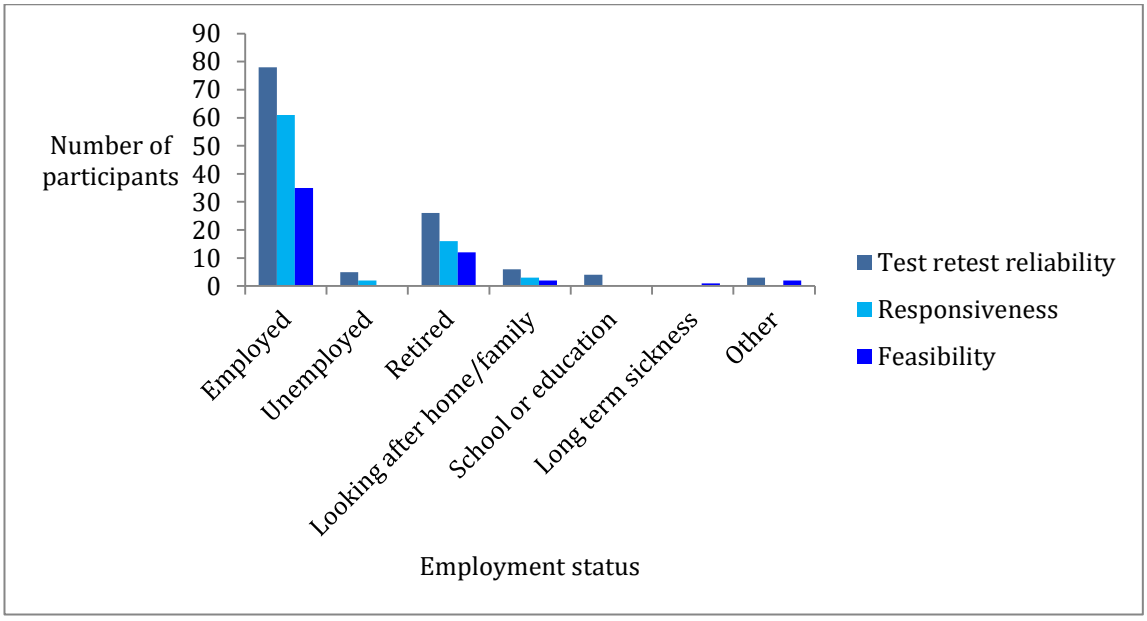


Figure 5.13. Employment status reported by participants

Ethnicity

Patients' disclosure of ethnicity was an optional question. A total of 80.5% of patients described themselves as "white British/Irish" and 9.7% as "white other". "Asian/Asian British" and "Black/African/Caribbean/Black British" were the ethnic groups described by 2.7% of the participants respectively, with 2% of participants describing themselves as "Mixed/multiple ethnic groups", and 2.4% as "other ethnic groups". The numbers of participants in each ethnic group are shown in Figure 5.14.

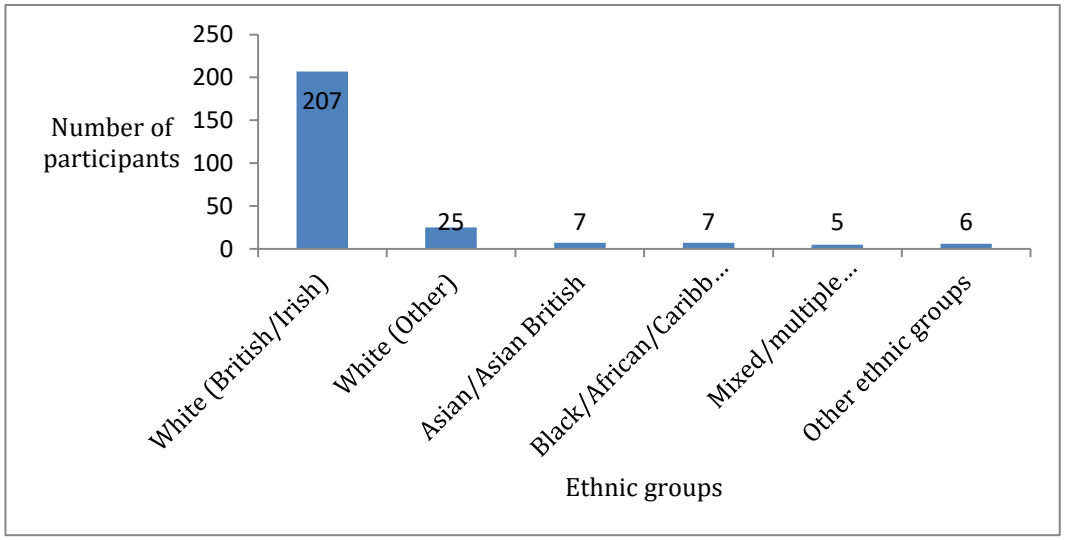


Figure 5.14. Participants' self-reported ethnicity

Sex

All participants provided data concerning their sex. A total of 59.9% (n=154) of participants were female, and 40.1% (n=103) were male as shown in Figure 5.15.

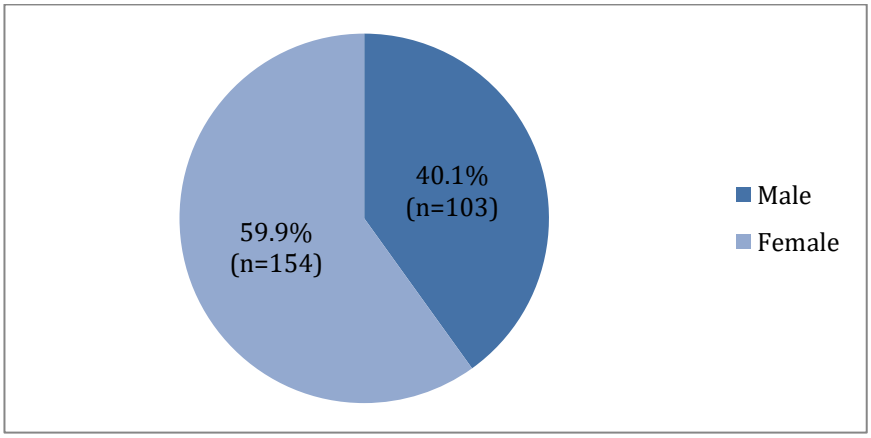


Figure 5.15 Sex of study participants

The duration of current symptoms

Patients reported a range of symptoms in accordance with recruitment for different arms of the study *i.e.* the responsiveness strand recruited participants with symptoms of six weeks or less duration, the test-retest strand recruited patients with stable, unchanging symptoms of 13 weeks or more, and the feasibility recruited patients with symptoms from 0-13 weeks' duration. The data concerning duration of symptoms are described as acute (0-6 weeks), subacute (7-12 weeks), and chronic (13 weeks or more).

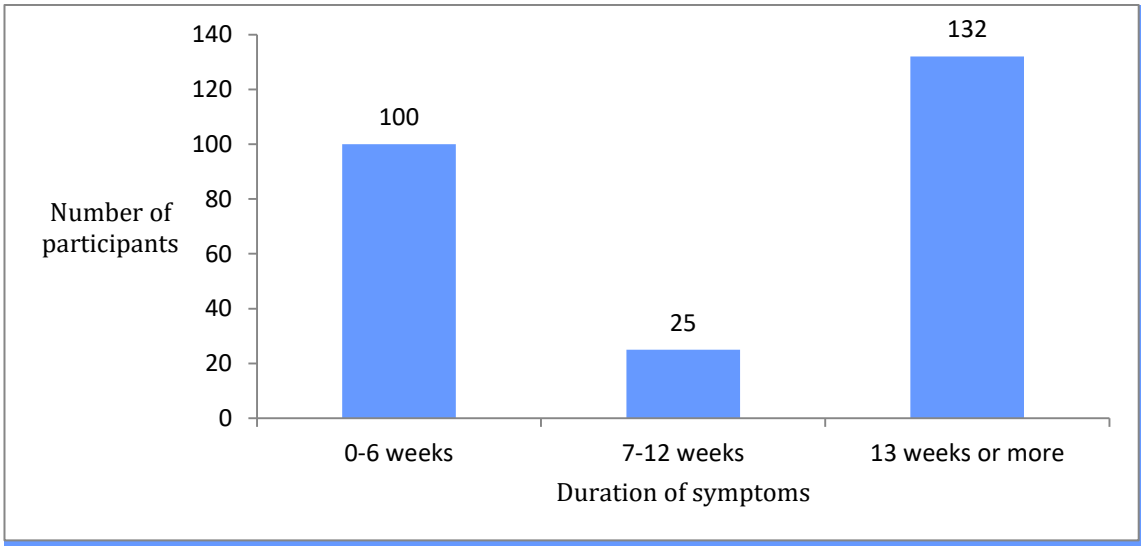


Figure 5.16. Duration of participants' symptoms.

Figure 5.16 describes the patients reporting symptoms of 0-6 weeks duration and represents 38.9% of the participant sample, 9.7% of participants reported symptoms lasting from 7-12 weeks, and 51.4% reported symptoms of 13 weeks or more.

Participants' general health status

Participants were asked to describe their general health status. A total of 37.7% reported their general health as "very good", 50.2% as "good", 11.2% as "fair", and 0.8% as "bad". No participants reported their health as "very bad". The data for all participants are presented in Figure 5.17.

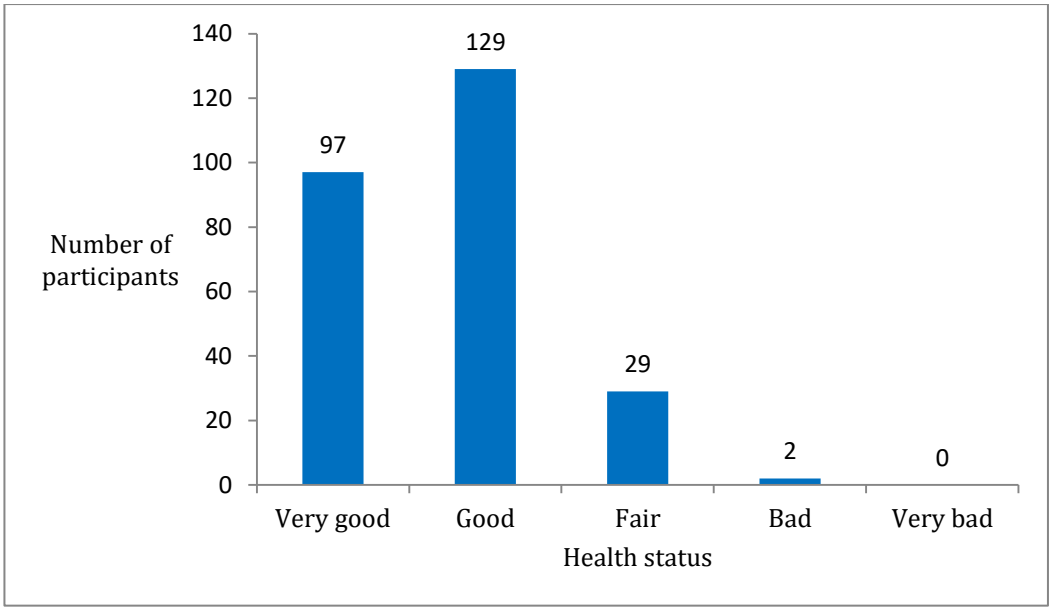


Figure 5.17. Participants' general health status

Waiting time to first appointment offered

Participants were asked to record the time until the first appointment they were offered by a practice on making contact. This service data is presented in Figure 5.18.

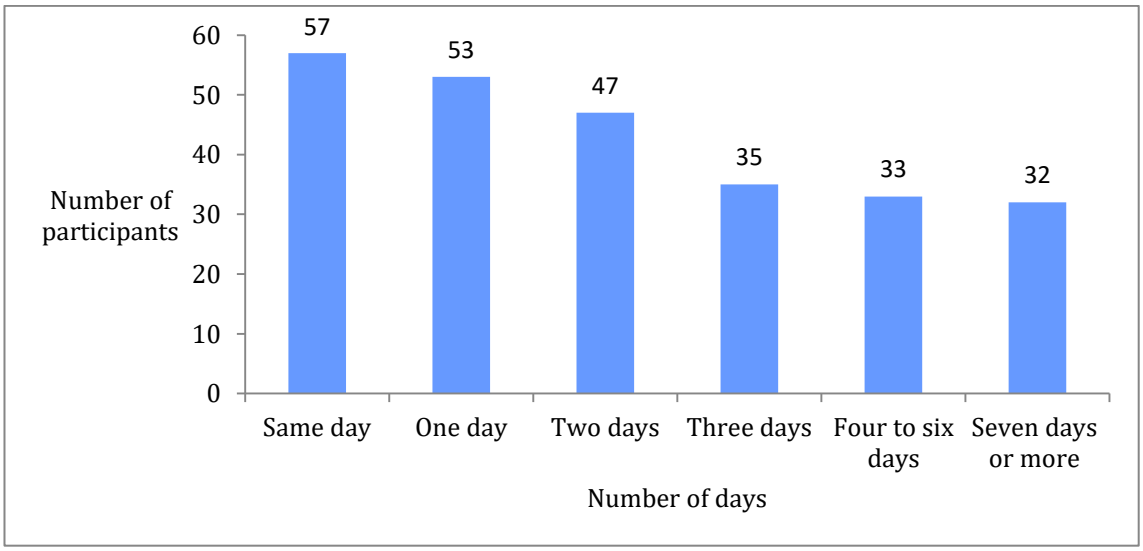


Figure 5.18. Waiting time until the first appointment offered by a practice

A total of 22.2% of participants reported they were offered an appointment the same day they made contact with a practice, 20.6% waited one day, 18.3% waited two days, 13.6% waited three days, 12.8% waited four to six days, and 12.5% reported waiting seven days or more.

The main reason for a participant seeking treatment

This information was collected qualitatively using a text box within the PROM app. Patients could report single or multiple reasons so the final total of contributions exceeds 100%.

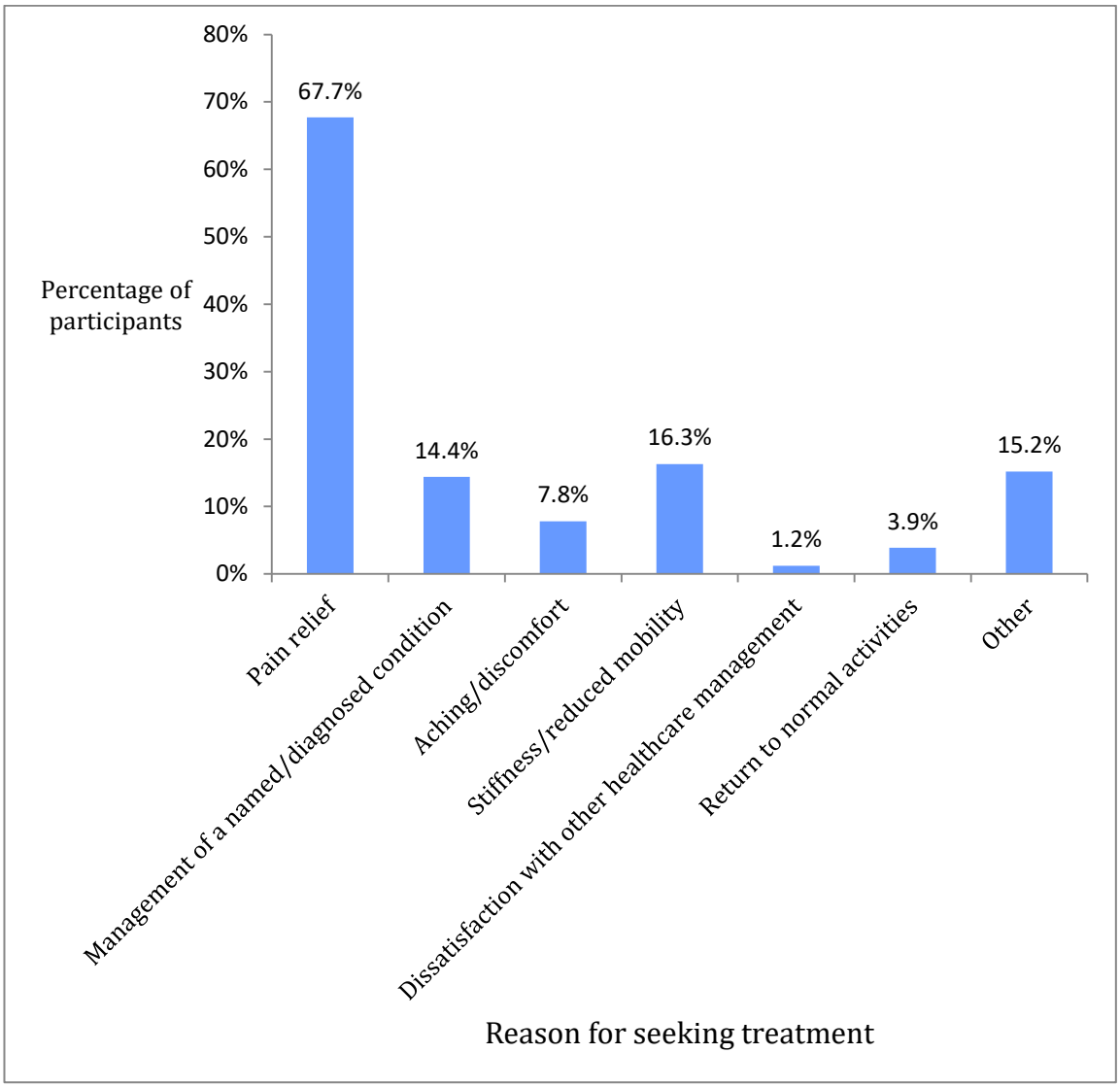


Figure 5.19. Participants' main reasons for seeking treatment

The information contributed by patients and shown in Figure 5.19 indicated that pain relief was the most frequently-cited reason (67.7%) for seeking treatment in this participant sample. Additional reasons reported are companion symptoms including stiffness and discomfort. For other patients, they required help with a named or diagnosed disorder, and a small percentage (1.2%) cited dissatisfaction with other forms of healthcare management. Among the reasons summarised as

"other" were "prevention", "misalignment", "twisted pelvis", "maintenance", "investigation and reassurance", "locked back", and "wanting to get to the root of the problem".

The main areas affected by symptoms

Participants were asked to state their area of primary symptoms, and other areas of companion symptoms. The main symptom areas reported in Figure 5.20 are low back (58.8%), neck (23%), head (6.6%), and upper back (5.8%). Other body sites were reported including anterior thorax and peripheral joints in the remaining 5.8% of participants.

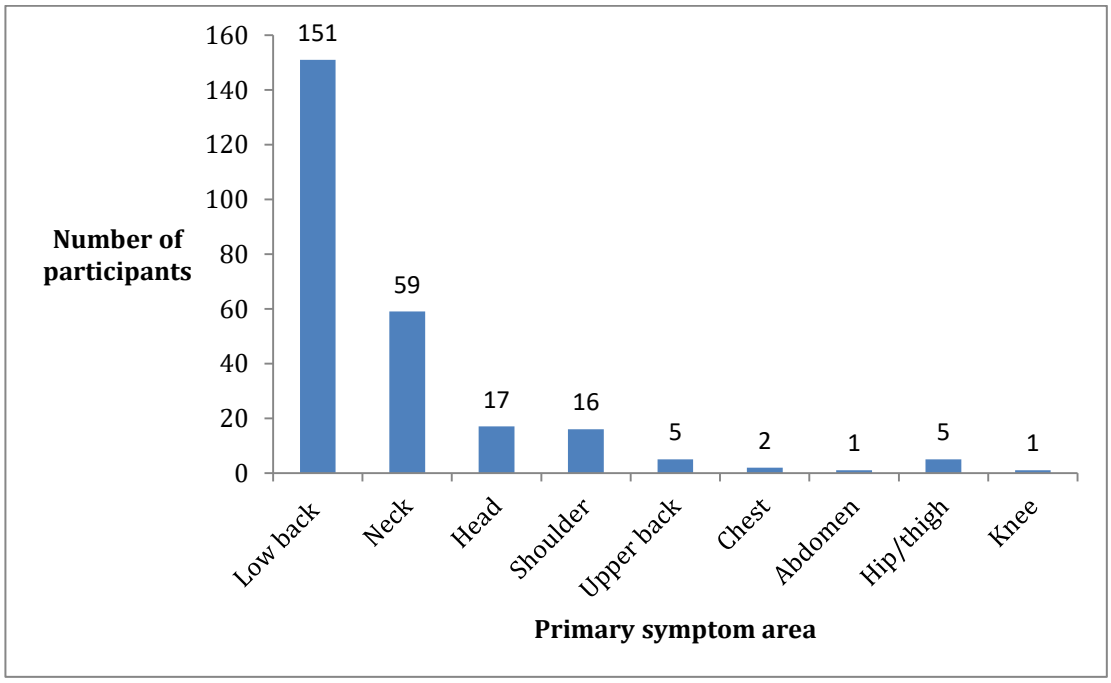


Figure 5.20. Primary symptom area reported by patients

The number of symptom areas reported by patients

Although participants were asked to state their primary area of symptoms, they were asked also to describe other symptom areas. Analysis of the data (shown in Figure 5.21) indicated that 51.4% of patients reported one symptom only, 17.9% reported two symptom areas, 15.2% reported three areas, 7.8% reported four areas, and 7.8% reported five or more symptom areas. Nine symptom areas were reported by 0.8% of patients. The mean number of symptoms reported was 2.11±1.53.

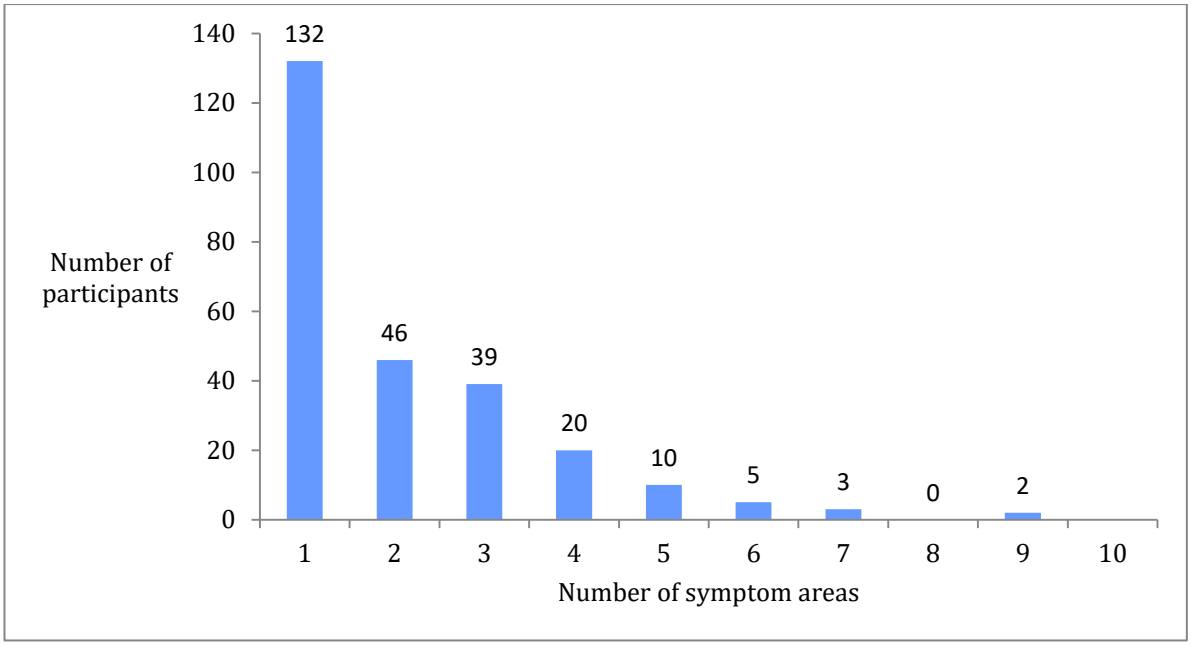


Figure 5.21. The number of symptom areas reported by patients.

Data were collected from participants using three Patient Reported Outcome Measures (PROMs). The data from each individual PROM will be described in turn, and further analysis will be shown relating to the measurement properties of the PROMs when used in an electronic format and in an osteopathic private practice setting.

5.5.2 Outcome data

The Bournemouth Questionnaire (BQ)

Data were collected using the BQ in each of the strands of the pilot. Baseline and one week data were collected for all three strands (test-retest reliability, feasibility, and responsiveness), and six week data were collected for the feasibility and responsiveness strands. Analysis of responses to individual questions from the BQ is shown in Figure 5.22. The scores from each of the individual domains are summed to produce a single score *i.e.* the sum score. The mean of the sum scores has been calculated for each strand of the PROM pilot and these are shown in Table 5.6.

Table 5.6. Scores for the BQ in each strand of the PROM pilot

	Responsiveness	Test-retest reliability	Feasibility
Baseline (n=83)	n=83 Mean = 32.88 ±13.17 Range: 2-63 Median: 27	n=122 Mean=20.98±7.1 Range: 1-55 Median: 20	n= 52 Mean = 19.08±6.7 Range: 2-47 Median: 19
One week	n=28 Mean = 12.29±17.92 Range: 0-46 Median: 13	n=46 Mean=12.93±4.3 Range:0-55 Median: 7	n= 25 Mean = 10.08±3.2 Range: 0-32 Median: 9
Six weeks	n=24 Mean = 10.25±10.63 Range: 0-37 Median: 6		n= 24 Mean = 11.21±3.1 Range: 0-32 Median: 6

When reviewing Figure 5.22, the legend refers to:

1. Over the past few days, on average, how would you rate your pain on a scale where 0 is no pain, and 10 is worst pain possible?

2. Over the past few days, on average, how has this complaint interfered with your daily activities (housework, washing, dressing, lifting, walking, reading, driving, climbing stairs, getting in/out of bed/chair, sleeping) on a scale where 0 is no interference and 10 is completely unable to carry on with normal activities?

3. Over the past few days, on average, how has this complaint interfered with your normal social routine including recreational, social and family activities, on a scale where 0 is no interference and 10 is completely unable to participate in and social and recreational activity?

4. Over the past few days, on average, how anxious (uptight, tense, irritable, difficulty in relaxing/concentrating) have you been feeling, on a scale where 0 is not at all anxious and 10 is extremely anxious?

5. Over the past few days, how depressed (down-in-the dumps, sad, in low spirits, pessimistic, lethargic) have you been feeling, on a scale where 0 is not at all depressed and 10 is extremely depressed?

6. Over the past few days, how do you think your work (both inside the home and/or employed work) has affected this complaint, on a scale where 0 is make it no worse and 10 is make it very much worse?

7. Over the past few days, on average, how much have you been able control (help/reduce) and cope with your pain on your own, on a scale where 0 is I can control it completely and 10 is I have no control whatsoever?

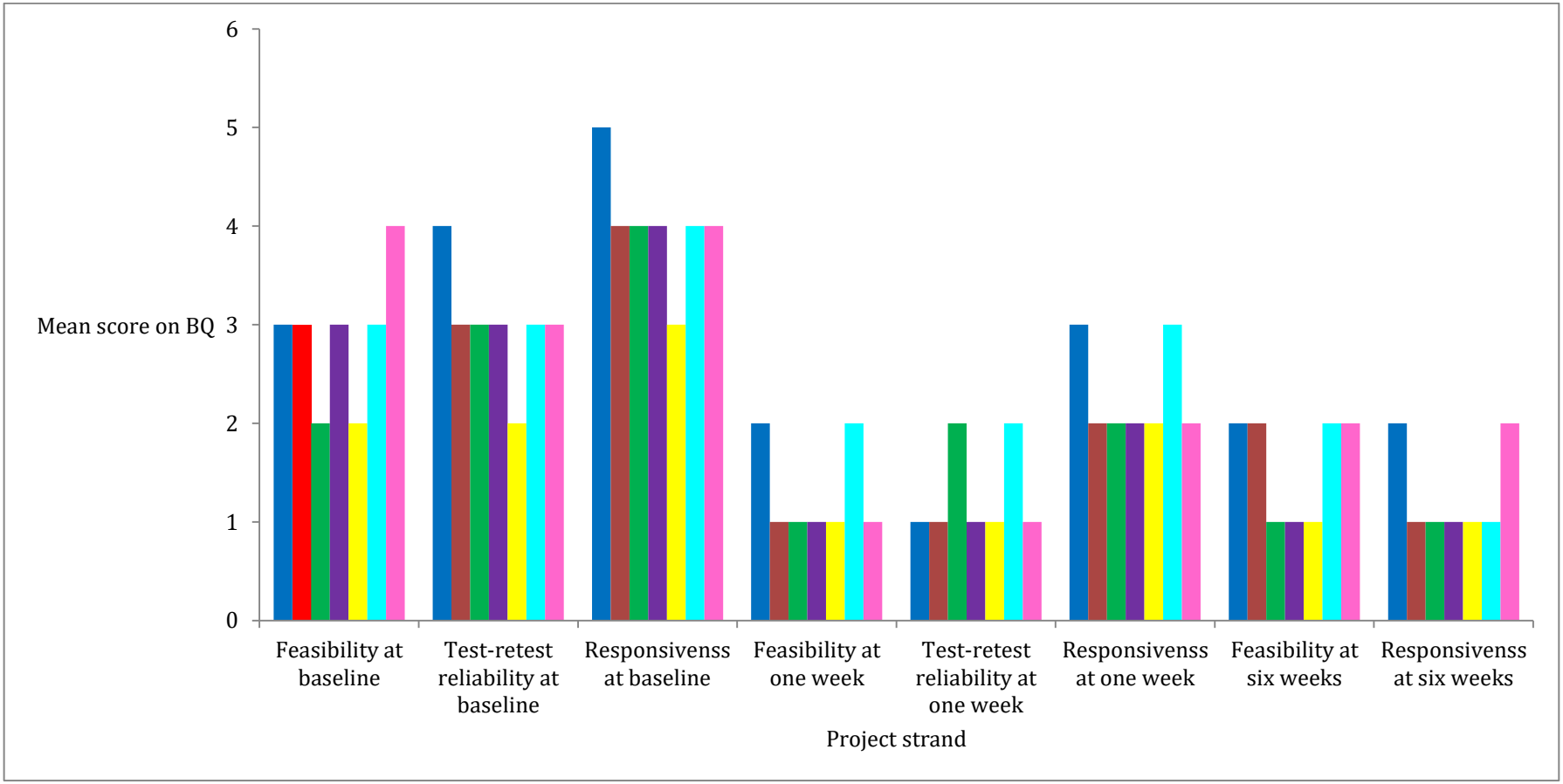


Figure 5.22. Data for each question on the BQ scale for all strands of the pilot project*

Responsiveness data

The mean percentage change score for the BQ questionnaire was 58.94%, calculated using.

$$\text{Percentage change} = \frac{\text{baseline – discharge scores}}{\text{Baseline}} \times 100$$

Mean improvement in score was 12.31, and mean deterioration was 6. The mean value for BQ scores at baseline in the responsiveness strand was 32.88±13.17, and at six weeks was 10.25±10.63.

Among the participants who contributed BQ data at more than one time point, 18.75% noted deterioration in their symptoms between baseline and week one; this reduced to 8.3% at the end of week six. The change for each patient at baseline, one week, and six weeks is presented visually in Appendix 5.4

The BQ scores for the responsiveness strand at baseline are shown in Figure 5.23.

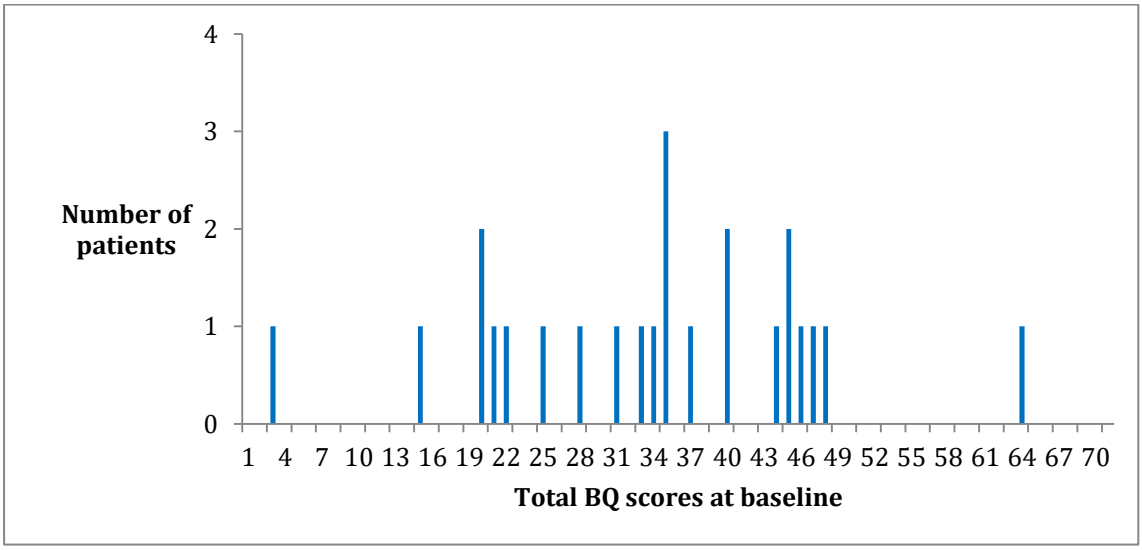


Figure 5.23 Total BQ scores for patients at baseline in the responsiveness strand

The scores for six weeks post treatment in the responsiveness strand are shown in Figure 5.24.

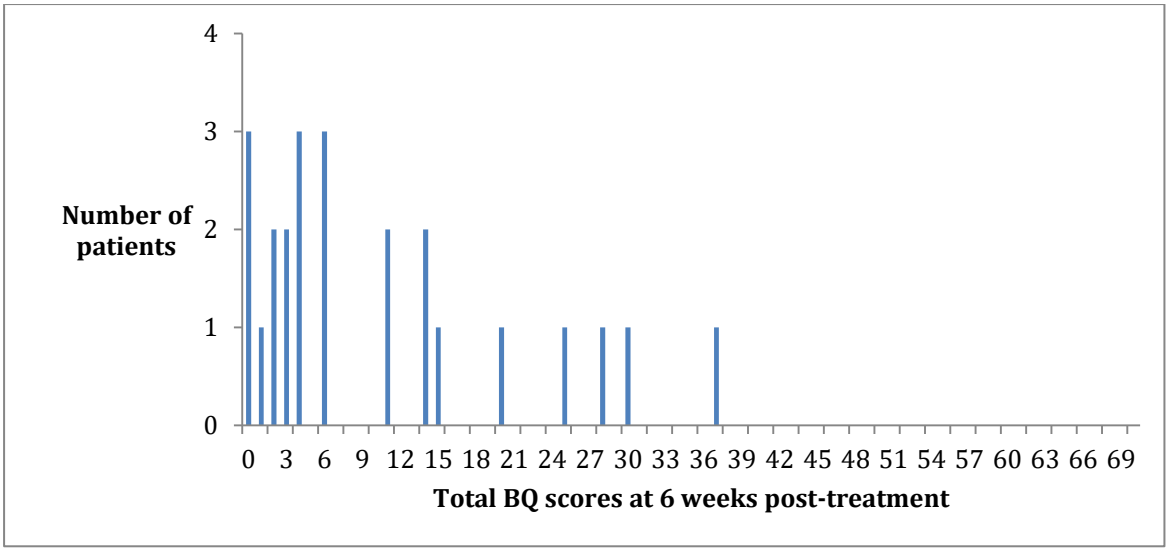


Figure 5.24 Total BQ scores for patients at 6 weeks post-treatment in the responsiveness strand

The change in BQ score data for all patients providing follow up data in the responsiveness strand has been summarised in a boxplot as shown in Figure 5.25.

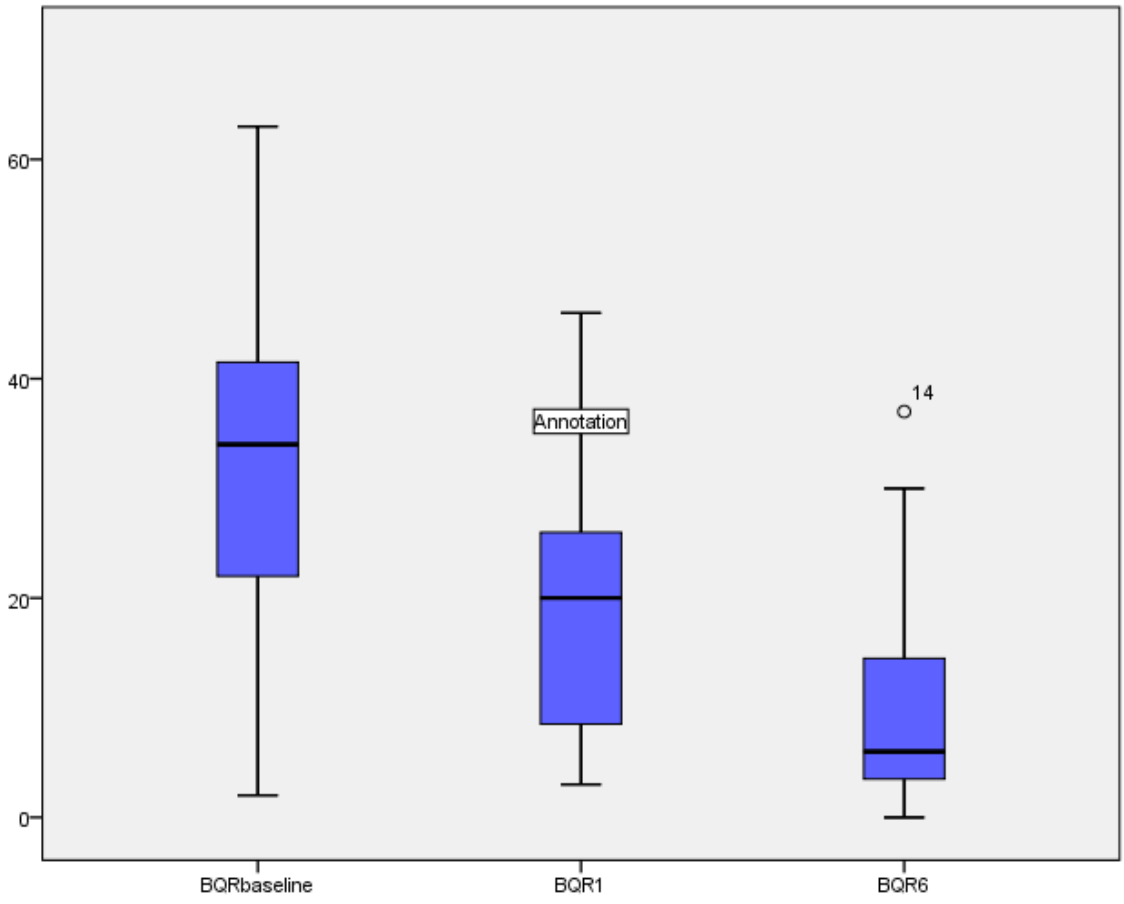


Figure 5.25. Box plot of change in BQ scores in responsiveness strand.

For participants in the test-retest strand, their baseline scores are shown in Figure 5.26.

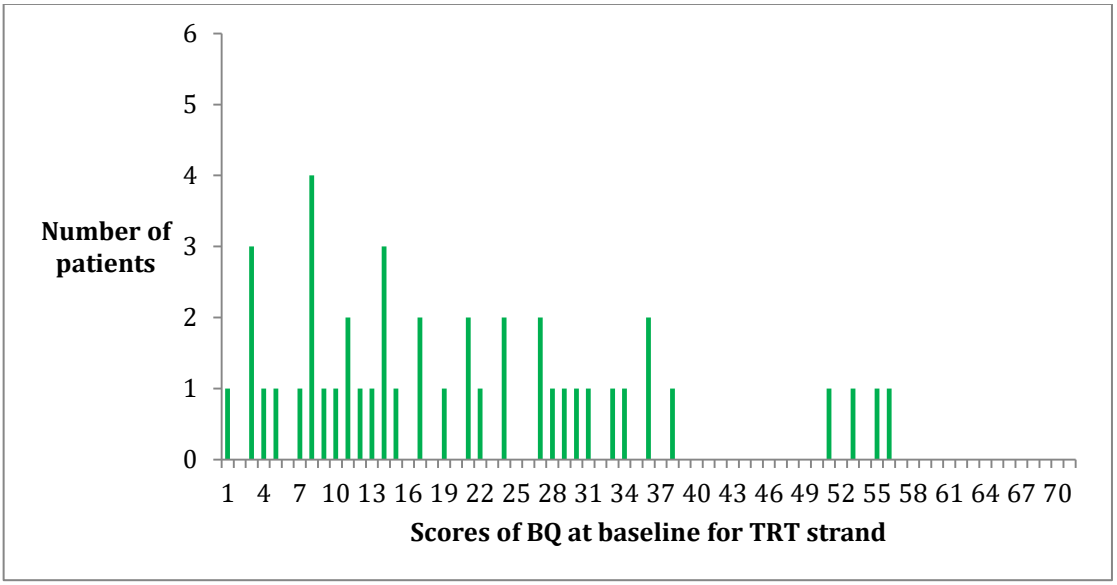


Figure 5.26 Baseline BQ scores for participants in the TRT strand

Follow up data at one week post treatment for participants in the TRT strand are shown in Figure 5.27.

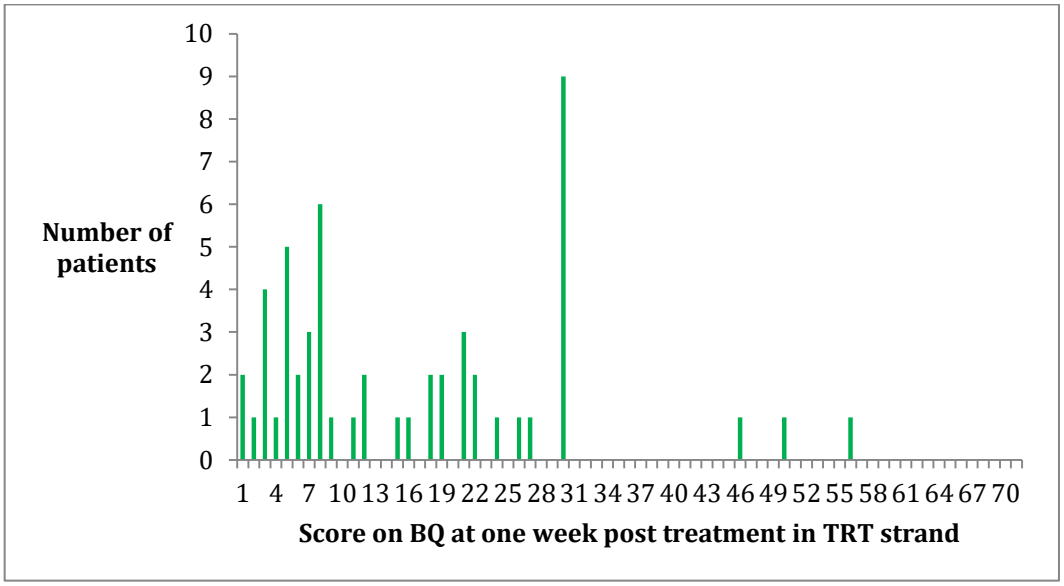


Figure 5.27 Scores in the BQ for participants in the TRT at one week post treatment

For patients in the feasibility strand, their baseline scores are shown in Figure 5.28.

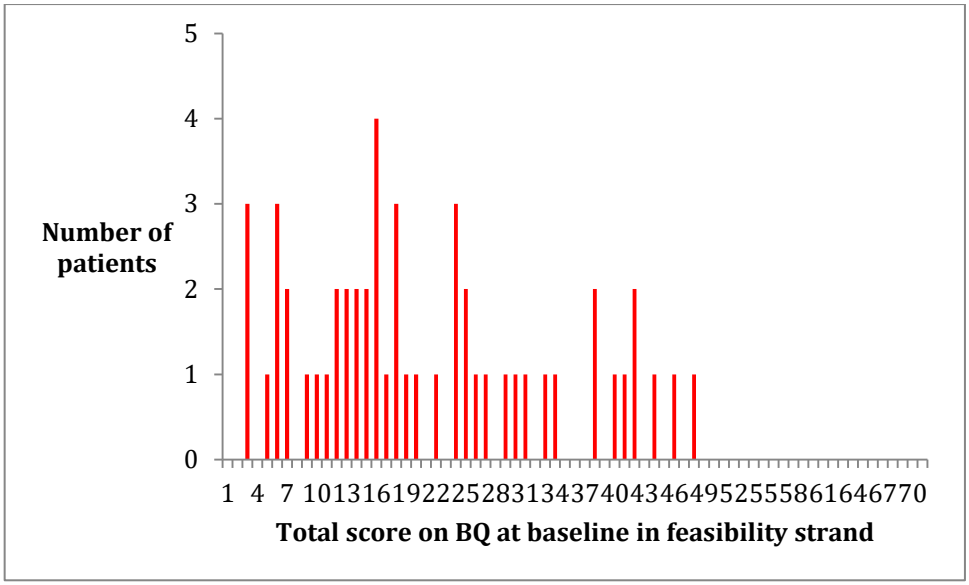


Figure 5.28 Baseline BQ scores for participants in the feasibility strand

Follow up data at six weeks for these patients are shown in Figure 5.29.

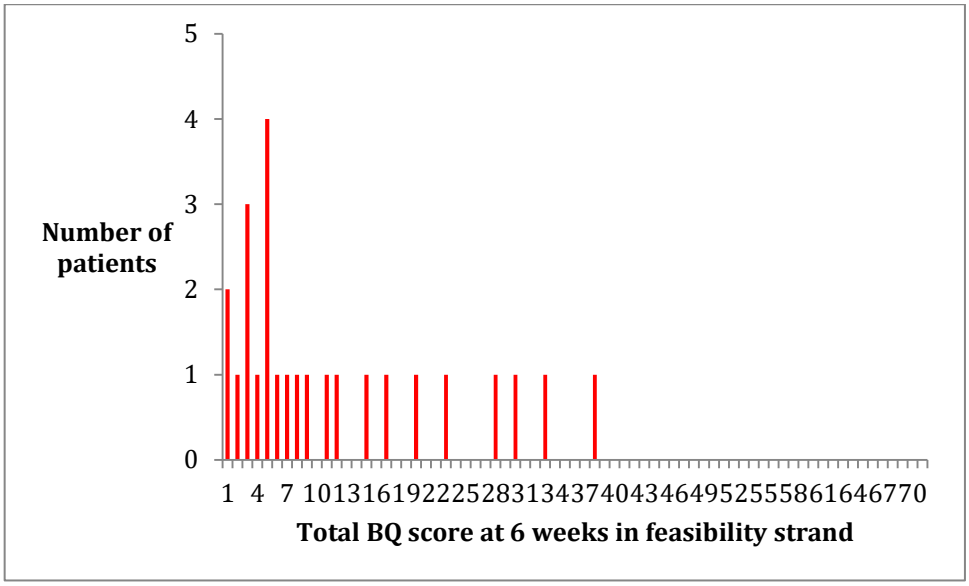


Figure 5.29 Scores in the BQ for participants in the feasibility strand at six weeks post-treatment

The Roland Morris Disability Questionnaire (RMDQ)

Data were collected using the 24-item version of the RMDQ at baseline, one week, and six weeks for each strand. These are presented in Table 5.7. Individual participants' scores on the RMDQ at each time point are shown in Appendix 5.5.

Table 5.7 RMDQ data for time points in all strands of the PROM pilot

	Responsiveness	Test-retest reliability	Feasibility
Baseline	n= 79 Mean= 6.13 Range: 0-20 Median: 5	n=110 Mean= 4.83 Range: 0-19 Median: 4	n= 43 Mean= 4.26 Range: 1-15 Median: 3
One week	n= 26 Mean= 4.08 Range: 0-12 Median: 3	n=34 Mean= 3.38 Range: 0-10 Median: 2	n= 16 Mean= 2.5 Range: 1-7 Median: 2
Six weeks	n= 12 Mean= 4.08 Range: 1-8 Median: 4		n= 17 Mean= 3.29 Range: 1-10 Median: 2

Scores on the RMDQ at baseline in the responsiveness strand are shown in Figure 5.30.

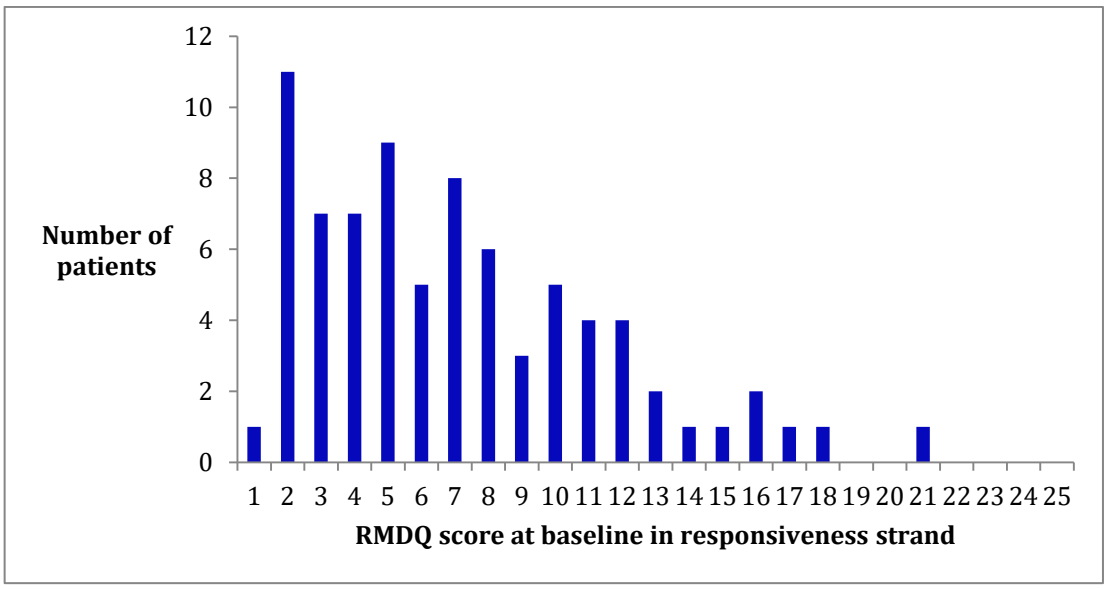


Figure 5.30. Total scores for the RMDQ in the responsiveness strand at baseline.

Scores on the RMDQ at six weeks post treatment are shown in Figure 5.31.

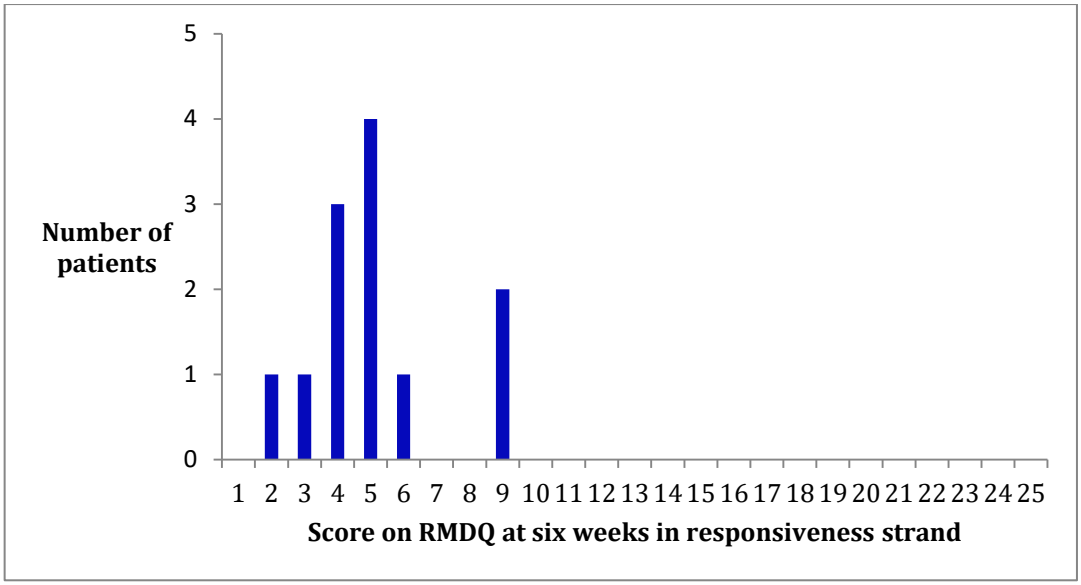


Figure 5.31. Total scores for the RMDQ in the responsiveness strand at six weeks.

These data have been summarised in a boxplot as shown in Figure 5.32.

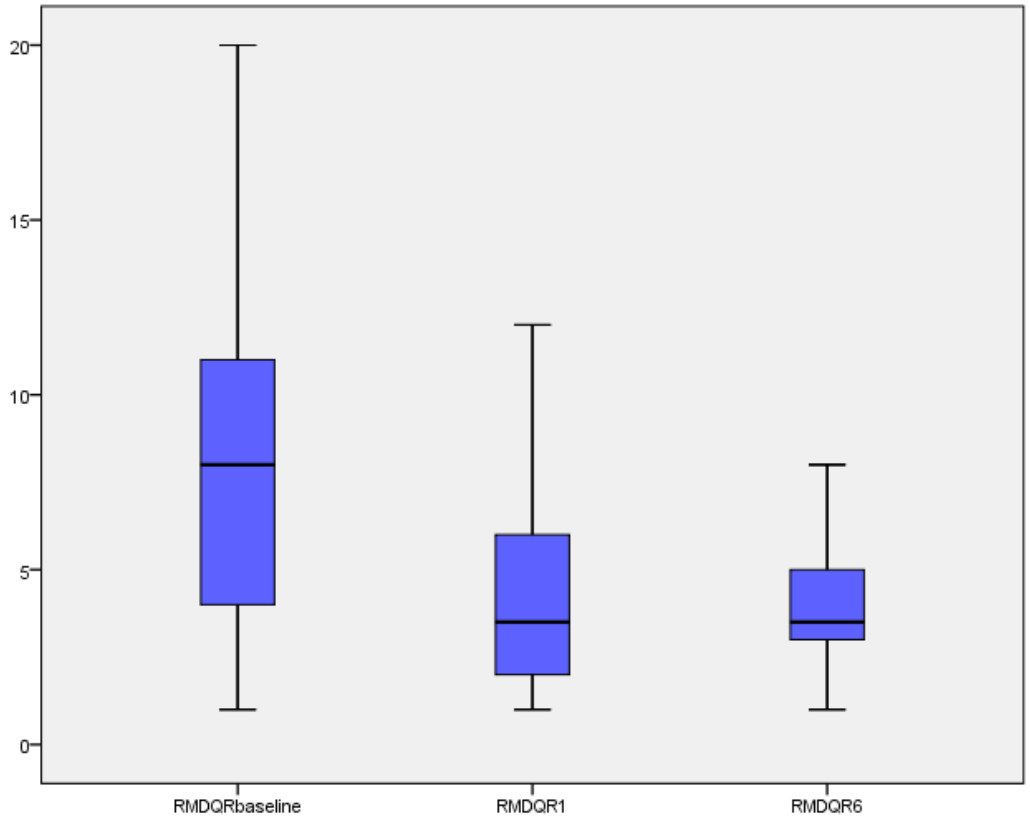


Figure 5.32. Box plot of change in RMDQ scores in responsiveness strand.

In the remaining strand of the project, the TRT, RMDQ data are shown at baseline in Figure 5.33.

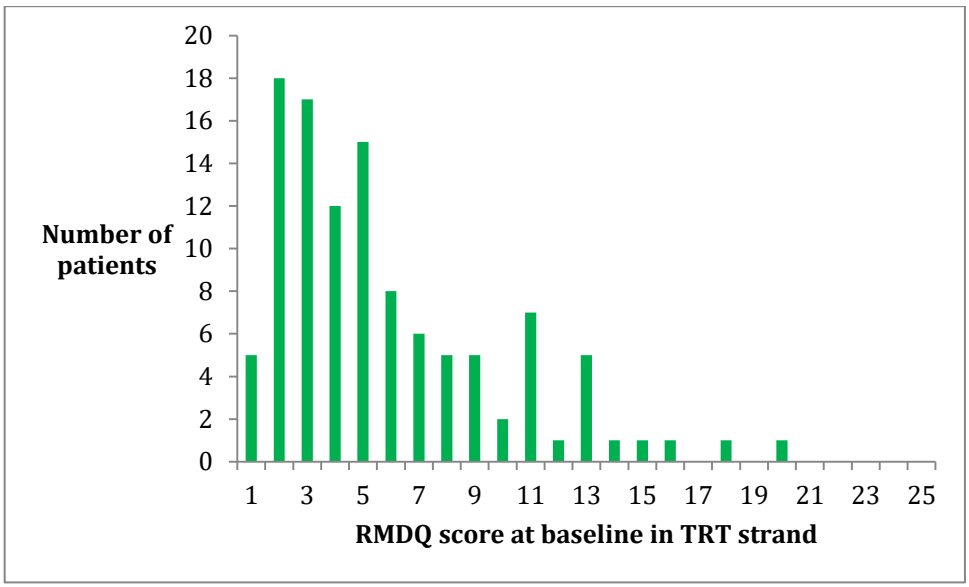


Figure 5.33. Total scores for the RMDQ in the test-retest strand at baseline.

Scores on the RMDQ at one week post treatment are shown in Figure 5.34.

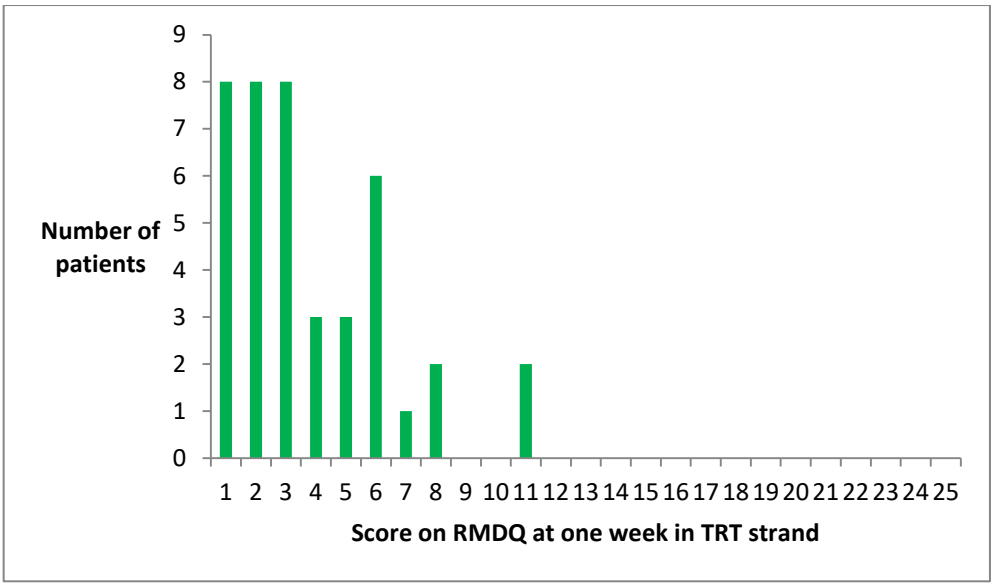


Figure 5.34. Total scores for the RMDQ in the test-retest strand at one week.

For patients in the feasibility strand of the study, their RMDQ scores are shown in Figure 5.35.

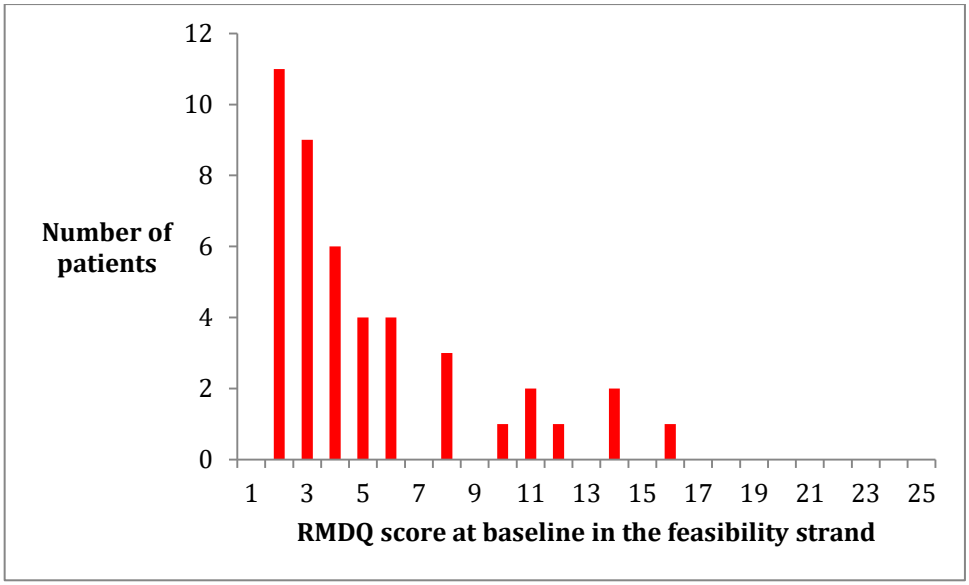


Figure 5.35. Total scores for the RMDQ in the feasibility strand at baseline.

Scores on the RMDQ at six weeks post treatment are shown in Figure 5.36.

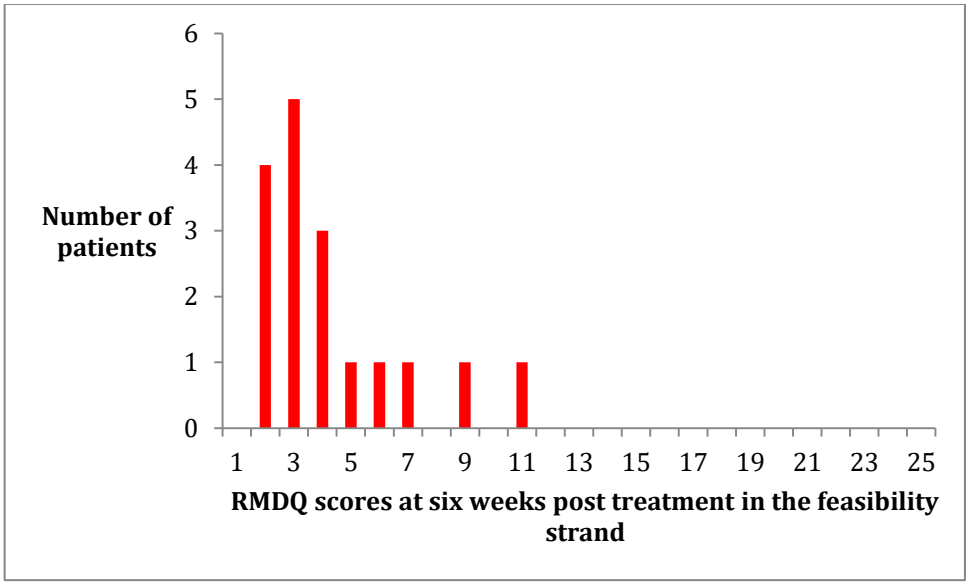


Figure 5.36. Total scores for the RMDQ in the feasibility strand at six weeks.

The progressive change in scores for all time points and in all participants in the responsiveness strand in the RMDQ is shown in Appendix 5. Negative scores indicate deterioration in symptoms. Steady improvement in symptoms between baseline and six weeks were reported by 74.2% of patients, and 5.3% reported a steady deterioration in symptoms over the same time period. A total of 10.5% of patients reported a deterioration between baseline and week one followed by an overall

improvement, 5.3% reported an improvement from baseline to week one followed by an overall deterioration, and 5.3% an improvement from baseline to week one, but symptoms increased after this initial improvement between week one and week six but this still represented an overall improvement from baseline score on the RMDQ.

A total of 12 complete datasets for the RMDQ were gathered. This included nine patients who reported individual improvements at each time point. There was an increase in symptoms for three patients who symptoms increased by 4-, 1-, and 2-points respectively.

The mean percentage change score for the RMDQ questionnaire was 27.59%.
Mean change in patients reporting improvement = 7.11
Mean deterioration = 2.33
The mean value for RMDQ scores at baseline in the responsiveness strand was 8.42±5.33, and at six weeks was 3.67±2.43.

Visual Analogue Scale (VAS)

Data were collected using a 100mm VAS at baseline, one week, and six weeks for each strand. These are presented in Table 5.8.

Table 5.8 VAS data for time points in all strands of the PROM pilot

	Responsiveness	Test-retest reliability	Feasibility
Baseline	n= 83 Mean= 47.39 Range: 6-90 Median: 49	n= 118 Mean= 35.82 Range: 0-88 Median: 34	n= 51 Mean= 35.35 Range: 1-71 Median: 30
One week	n= 26 Mean= 27.23 Range: 1-88 Median: 26	n= 42 Mean= 27.21 Range: 5-70 Median: 22	n= 23 Mean= 19.78 Range: 4-58 Median: 18
Six weeks	n= 20 Mean= 18.86 Range: 0-43 Median: 17		n= 23 Mean= 22.17 Range: 3-62 Median: 17

Participants submitted VAS scores (using a slider) at baseline, one week, and six weeks in the responsiveness strand of the study. The mean value of VAS in the responsiveness strand was 51.81±21.29 at baseline, and 18.06±13.36 at six weeks.
A total of 16 complete datasets for the VAS were gathered. This included 13 patients who reported individual improvements at each time point. There was an increase in symptoms at week one for two patients who symptoms increased by 1-, and 5-points respectively. For two patients, their symptoms decreased between baseline and one week but increased by 19- and 10- points between one week and six weeks although there was an overall improvement between baseline and six weeks.

The mean percentage change score for the VAS questionnaire was 64%. Mean change in patients reporting improvement = 33.31. Scores on the VAS at baseline in the responsiveness strand are shown in Figure 5.37.

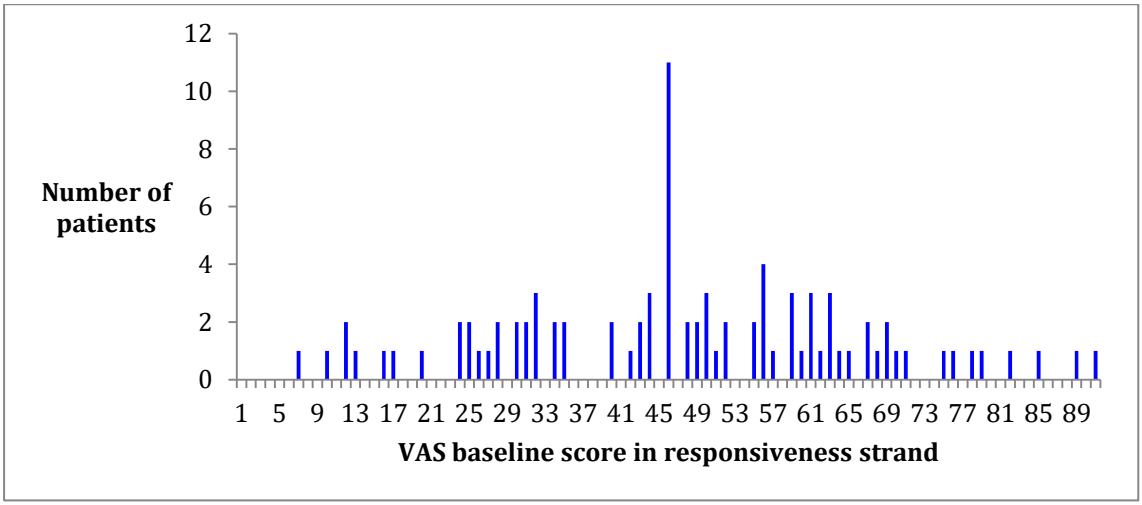


Figure 5.37. Total scores for the VAS in the responsiveness strand at baseline.

The VAS scores at six weeks post treatment are shown in Figure 5.38.

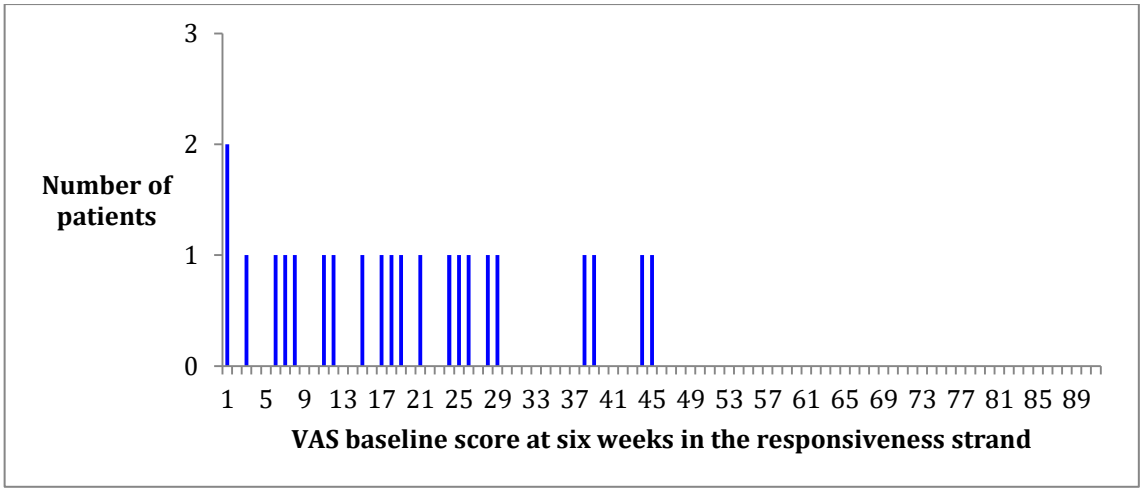


Figure 5.38. Total scores for the VAS in the responsiveness strand at six weeks.

These data have been summarised in a boxplot as shown in Figure 5.39.

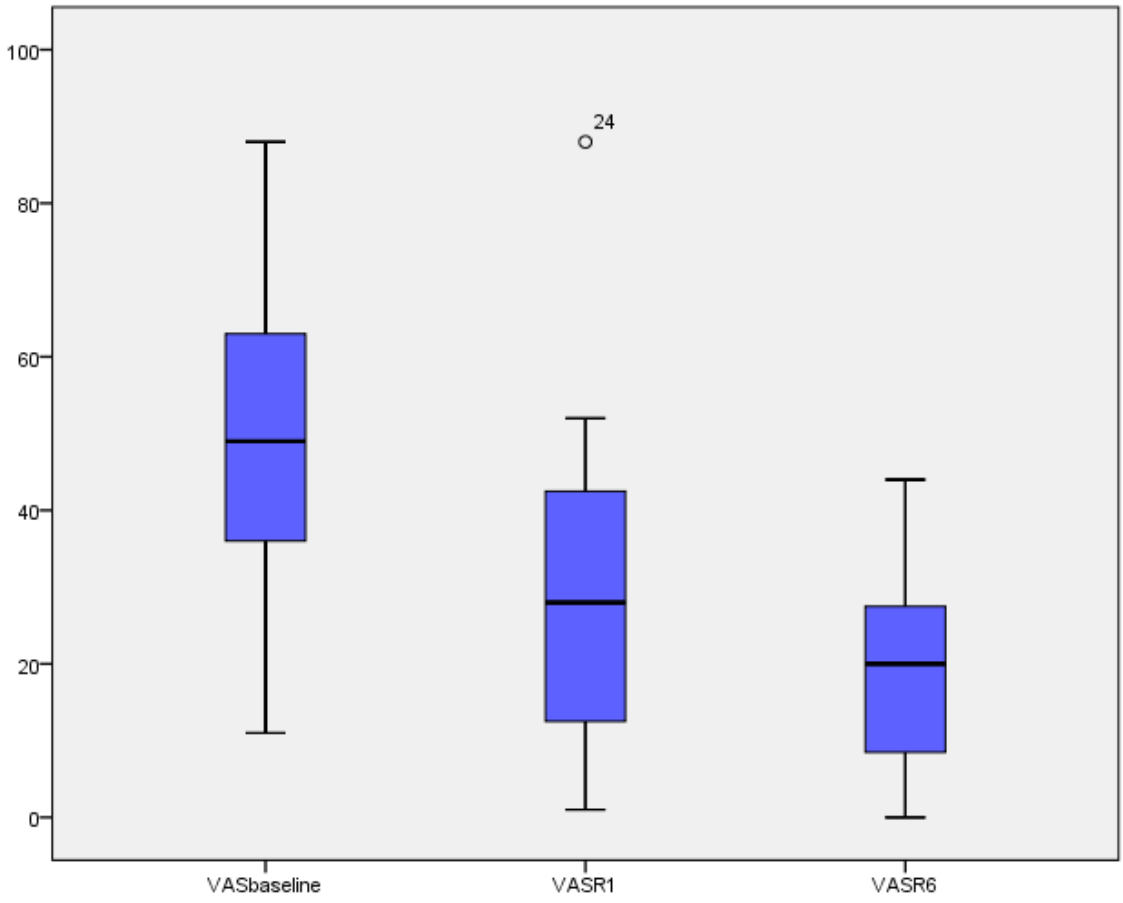


Figure 5.39. Box plot of change in VAS responsiveness scores.

For patients in the test-retest reliability (TRT) strand of the study, VAS scores at baseline are shown in Figure 5.40.

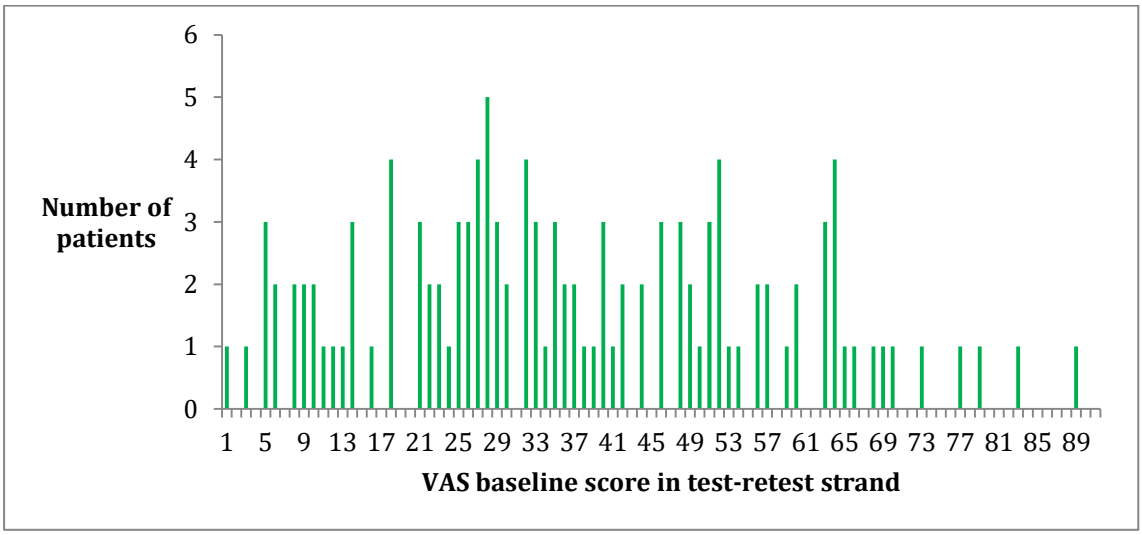


Figure 5.40. Total scores for the VAS in the test-retest strand at baseline.

The VAS scores in the TRT strand at one week post treatment are shown in Figure 5.41.

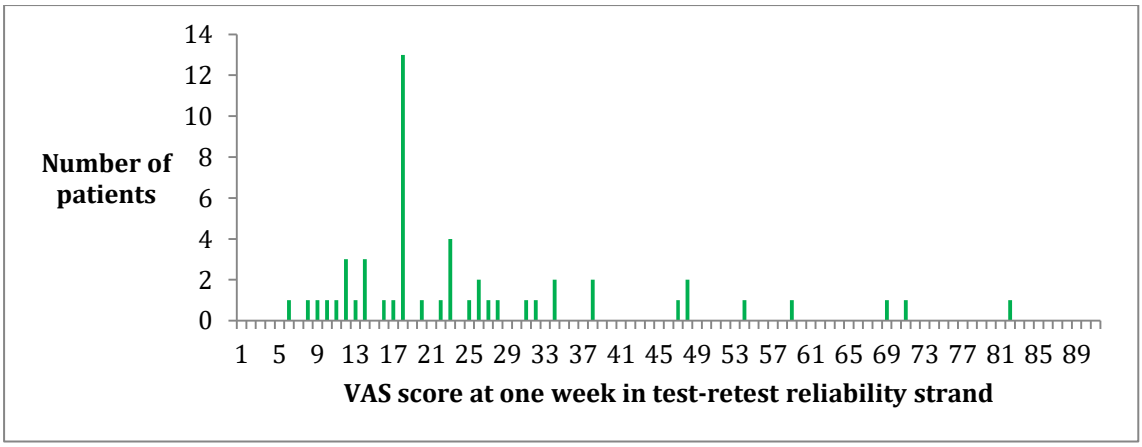


Figure 5.41. Total scores for the VAS in the test-retest strand at one week.

Patients in the feasibility strand contributed data at baseline and six weeks post treatment. These data are shown in Figures 5.42 and 5.43 respectively.

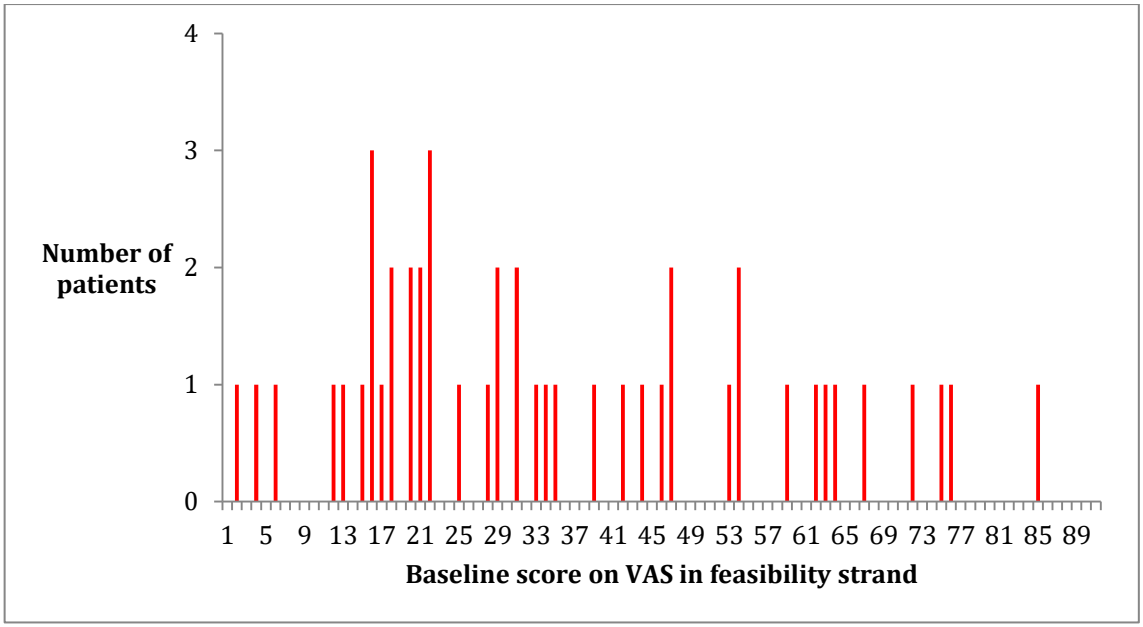


Figure 5.42. Total scores for the VAS in the feasibility strand at baseline.

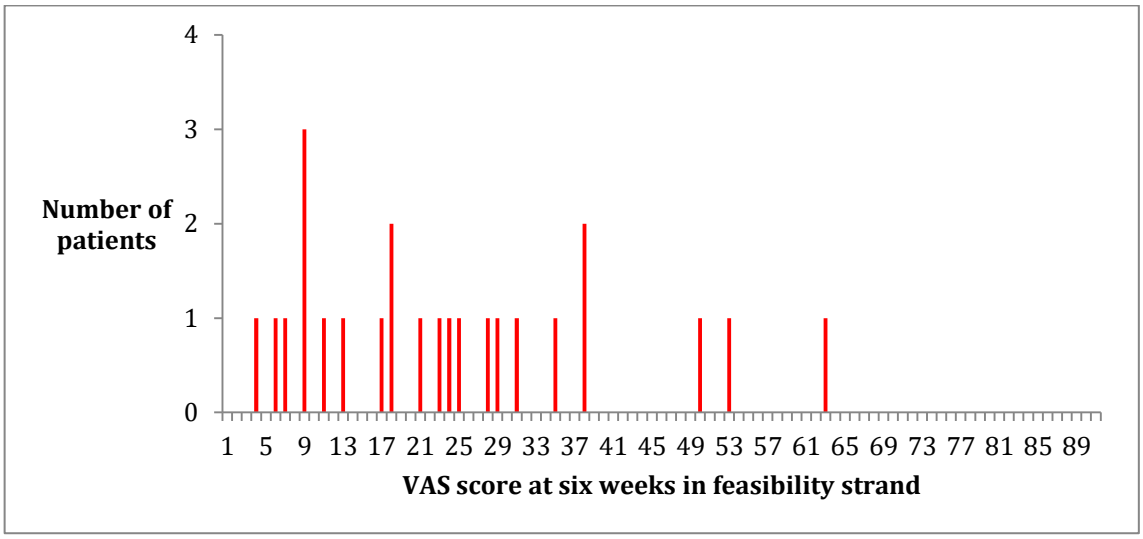


Figure 5.43. Total scores for the VAS in the feasibility strand at six weeks.

Transition question (TQ)

Figure 5.44 shows participants' responses to the enquiry concerning global change in symptoms. This is described for all strands of the study, and is measured at one week and six weeks post treatment for the feasibility and responsiveness strands, and one week only for the test-retest reliability strand.

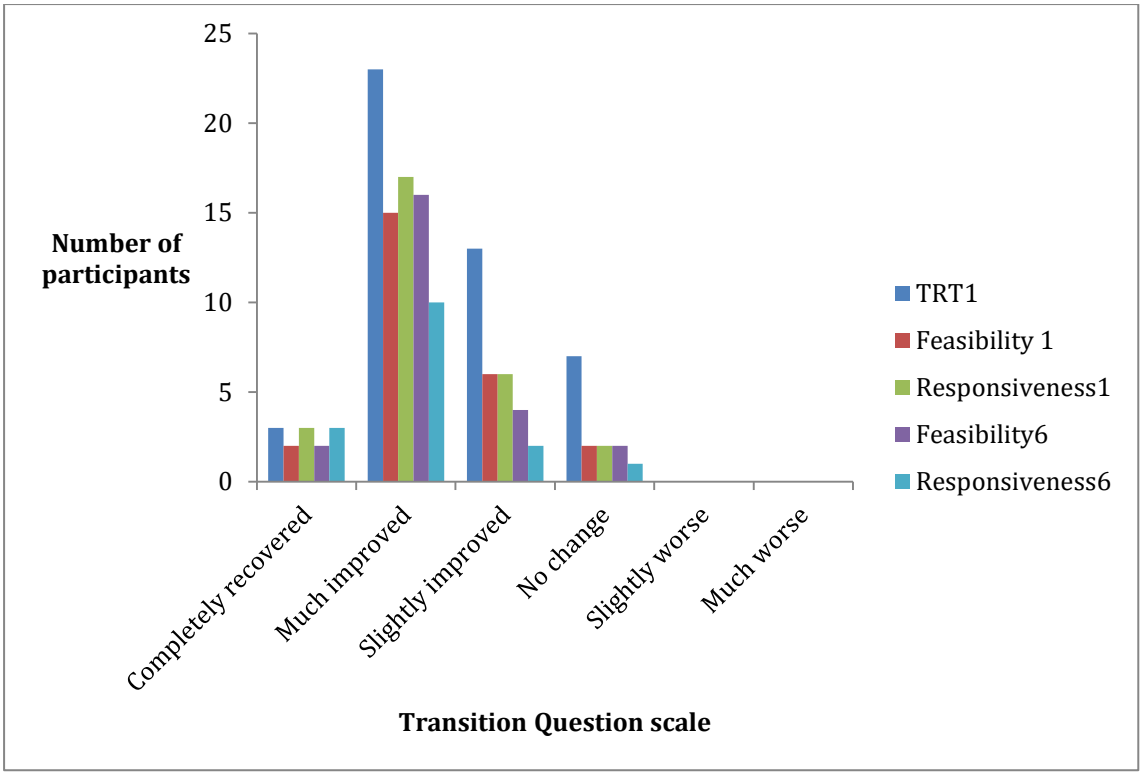


Figure 5.44. Scores on Transition Question scale for the test-retest reliability, feasibility, and responsiveness strands of the PROM pilot.

Secondary endpoints

While the descriptive data are information about the population involved in the study, and some of their characteristics, it does not allow any relationships to be studied

Patient satisfaction

Patient satisfaction was measured across all three strands of the study. This included measurement at one week post treatment for test-retest reliability, responsiveness, and feasibility, and at six weeks for responsiveness and feasibility. The findings for the individual strands are shown in Figure 5.45.

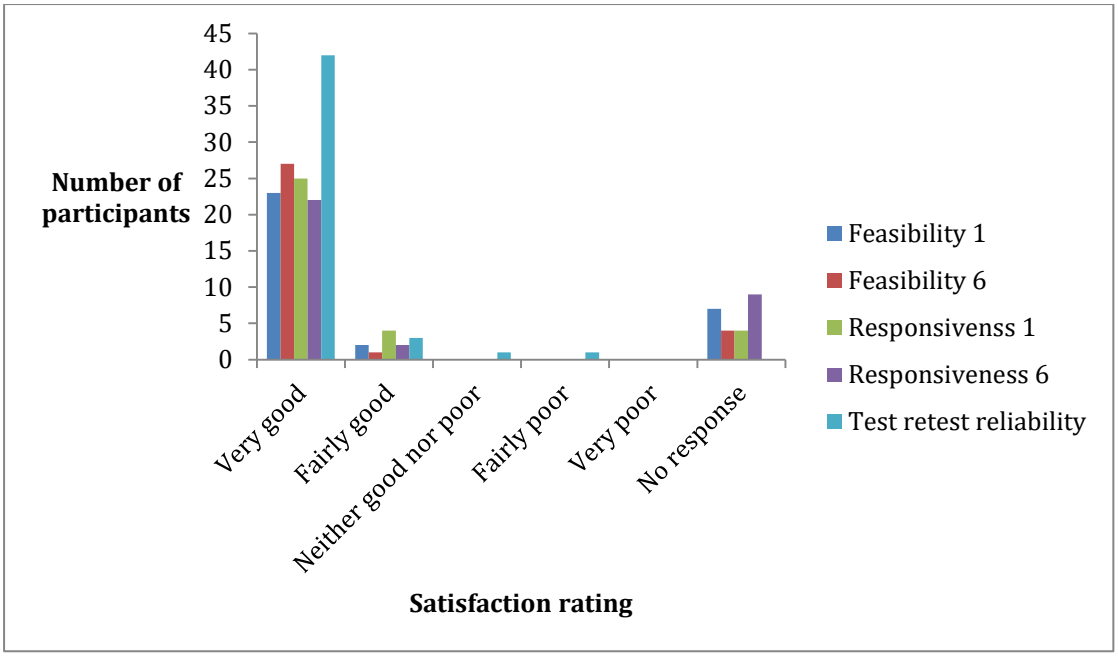


Figure 5.45. Patient satisfaction scores for test-retest reliability, feasibility, and responsiveness strands of the PROM pilot.

The relationship between patient satisfaction and other patient characteristics was explored.

Age vs satisfaction

When the relationship between age and satisfaction was explored at one week post treatment Chi square statistic 0.718 (df=12) in the responsiveness strand. At six weeks this value was 0.791 (df=12). This was mirrored in the test-retest reliability strand (0.622). A strong positive relationship was found between increasing age and satisfaction in both the responsiveness and test-retest reliability strands of the study. Relationships were also examined between sex and satisfaction, employment status and satisfaction, and ethnic status and satisfaction. No relationships were found on analysis of these variables using regression analysis.

Patient experience

Patient experience (PE) was measured for each strand of the study. For the test-retest strand patient experience is measured at one week post treatment, and for the feasibility and responsiveness strands it is measured at one week and six weeks post treatment. Scores for each strand of the PROM pilot are shown in Figure 5.46. PE was

explored to evaluate whether any relationship existed between PE and a patient's age, sex, ethnicity, or work status.

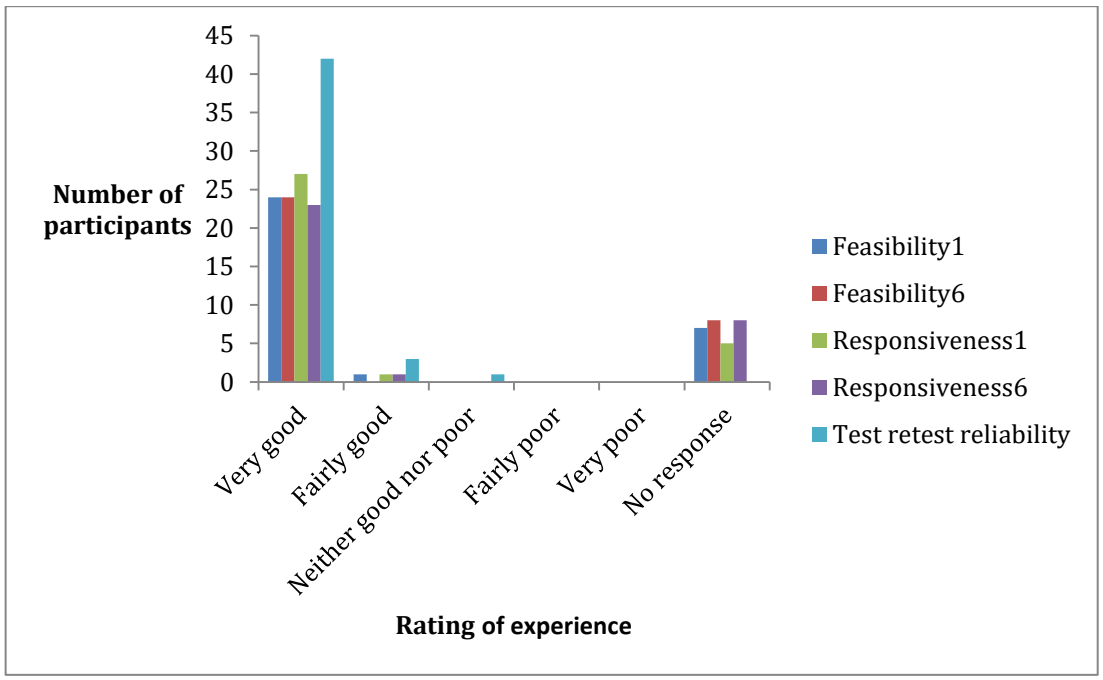


Figure 5.46. Patient experience scores for test-retest reliability, feasibility, and responsiveness strands of the PROM pilot.

Age vs experience

When age was compared with experience at one week post treatment, a Chi square statistic value of 0.495 (df=12) was obtained. At six weeks this value was 0.545 (df=6). Increasing age showed a positive relationship with increased experience of care. Relationships were explored between gender and experience, employment and experience, and ethnicity and experience. No relationships were found between these variables on analysis.

5.5.3 Evaluation of measurement properties

As mentioned in 5.3.2, the measurement properties of PROMs are important when considering their use in a range of clinical settings. The reliability and responsiveness of the BQ, the RMDQ, and the VAS will be explored in this section based on their use in an electronic medium and in an osteopathic setting. The interpretation of those findings will be discussed also.

Reliability

Although reliability can include inter-rater reliability, intra-rater reliability, and test-retest reliability, it is the latter with which I am concerned in this study. Statistical values for reliability were calculated for each of the Patient Reported Outcome Measure (PROM) used. Values for the Intra-class Correlation Coefficient (ICC) and Cronbach's α are described in turn for each PROM. The stability of the population in the test-retest reliability strand of the study was assessed based on the criteria that patients had experienced their symptoms for greater than 13 weeks; their symptoms were self-reported as unchanging; and they attended practices on a regular basis for "maintenance" treatment. The concept of repeated measurements to assess test-retest reliability is well established in the literature (Kirshner and Guyatt, 1985; McDowell, 2006). The interval between "test" and "retest" completion requires careful consideration. It must not be too long or there is the possibility symptoms may have changed, or too short and patients may complete the questionnaire remembering their first response rather than completing it de novo. Streiner and Norman report that although expert opinion varies in regard to an appropriate interval being from one hour to one year, they suggest an interval of 2 to 14 days is usual (Streiner and Norman, 2008). In the development of the Bournemouth Questionnaire, Bolton and Breen report that the initial questionnaire was administered to patients at their first visit prior to treatment. A second (retest) questionnaire was administered later the same day prior to any treatment taking place (Bolton and Breen, 1999). In other studies the baseline characteristics of patients have been gathered over an extended period of time while awaiting surgery. There has been an assumption of stability demonstrated through lack of change or minimal change during various measurement intervals (Kelly *et al.,* 2001) *e.g*. 106 days and 178 days (Stratford *et al.,* 2007; Kennedy *et al.,* 2005). Patients completed their test questionnaire prior to their osteopathic appointment, and their retest questionnaire one week later.

Bournemouth Questionnaire (BQ)

Analysis of the BQ test-retest data identified a value for the intraclass correlation coefficient (consistency) of 0.520 (95% CI 0.045 to 0.767), and a value for Cronbach's α of 0.791. The value for Pearson's correlation coefficient was 0.666 which was

significant at the 0.01 level. The SEM was calculated to be 2.79, and the SDC, a parameter of measurement error, to be 7.72.

Roland Morris Disability Questionnaire (RMDQ)

Analysis of the RMDQ test-retest data identified a value for the intraclass correlation coefficient (consistency) of 0.653 (95% CI 0.260 to 0.839), and a value for Cronbach's α of 0.844. The value for Pearson's correlation coefficient was 0.740 which was significant at the 0.01 level. The SEM was calculated to be 0.47, and the SDC to be 1.29.

Visual analogue scale (VAS)

Analysis of the VAS test-retest data identified a value for the intraclass correlation coefficient (consistency) of 0.599 (95% CI 0.013 to 0.838), and a value for Cronbach's α of 0.858. The value for Pearson's correlation coefficient was 0.763 which was significant at the 0.01 level. The SEM was calculated to be 2.73, and the SDC to be 7.56.

Responsiveness

A range of different values can be calculated for responsiveness using anchor – based and distribution-based methods. Values for individual change scores can be included to determine responsiveness, and have been described earlier in the results section. The area under the curve (AUC) has been calculated to evaluate responsiveness as this is regarded as an index of diagnostic discrimination, and in using a PROM we are attempting to detect change in the participants' reports of low back pain when managed with osteopathic care. TQ data are summarised as either "improved" or "not improved", and the percentages in each group at one week and six weeks are shown in Table 5.9.

Table 5.9 Responses to TQ for one week and six week follow up

	One week follow up	Summary of one week follow up	Six weeks follow up	Summary of six week follow up
Completely recovered	7.7% (n=2)	Improved 69.2% (n=18)	66.8% (n=16)	Improved 75.1% (n=18)
Much improved	61.5% (n=16)		8.3% (n=2)	
Slightly improved	23.1% (n=6)	Not improved 30.8% (n=8)	16.6% (n=4)	Not improved 24.9% (n=6)
No change	7.7% (n=2)		8.3% (n=2)	
Slightly worse	0		0	
Much worse	0		0	
Vastly worse	0		0	

Values for AUC were calculated for each of the PROMs included in the app. Values for changes at one week post-treatment and six weeks post-treatment are shown in Table 5.10.

Table 5.10 AUC values for all PROMs in the PROM app at one week and six weeks post-treatment

PROM	Time interval	AUC	95% Confidence interval
BQ	1 week	0.725	0.494 to 0.956
BQ	6 weeks	0.881	0.698 to 1.000
VAS	1 week	0.670	0.412 to 0.929
VAS	6 weeks	0.972	0.892 to 1.000
RMDQ	1 week	0.732	0.535 to 0.929
RMDQ	6 weeks[a]	0.778	0.398 to 1.000
RMDQ	6 weeks[b]	0.833	0.540 to 1.000

RMDQ[a] shows the value for the AUC with scores submitted with definite zeros.

RMDQ[b] shows the value for the AUC where zero values have been submitted where patients responded "none" to the question about symptoms at the end of the RMDQ which allowed the app to capture explicit data about the RMDQ clarifying whether patients had no symptoms captured by the RMDQ or whether the questionnaire was being ignored.

5.5.4 Interpreting change

As mentioned earlier, Jaeschke *et al.* defined the MIC 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 *et al*., 1989). The value for MIC refers to change related to the individual patient and not for groups of patients. Various methods have been proposed for calculating the value for MIC. The MIC is essentially a measure of interpretability of change. Both distribution and anchor-based approaches were described in Section 5.3.4, but this section will base MIC values on ROC curve data (Farrar *et al.,* 2001).

Bournemouth Questionnaire
MIC = 12.5 points

Roland Morris Disability Questionnaire
MIC = 4.5 points

Visual Analogue Scale
MIC = 27.5mm

While the values provide some useful data in isolation, it is important to compare the findings of this study to other literature, and place data in context with other measurement properties. This will be explored in the discussion section.

5.6 Discussion

The pilot of the PROM app has yielded a considerable amount of data. I will explore how, the descriptive and inferential data compare with other studies of a similar nature. I will then explore the data relating to the measurement properties and how these compare to data from other studies.

5.6.1 Comparison to existing studies

Although only a small number of studies exists which profile osteopathic care, they do provide descriptive data to which this study can be compared.

Descriptive data

The majority of studies have included retrospective data collection (Burton, 1981; Pringle and Tyreman, 1993; Hinkley and Drysdale, 1995; McIlwraith, 2003). One

more recent study undertook prospective data collection (Fawkes *et al.,* 2012). Age ranges in previous studies were similar with the largest group of patients being in age range 41-60 (44%) in Pringle and Tyreman, 1993; 47% in the 35-54 age range in Burton, 1981; 40-49 in McIlwraith, 2003, and 22% in 30-39 (Fawkes *et al.,* 2012). In this study, the largest group of patients were in the 50-59 age band across all project strands.

Gender has been discussed in all previous studies. In the PROM app pilot, 40.1% of males and 59.9% of males participated compared with 52% of males in Pringle and Tyreman; 50.2% of males in Burton; 63% males in McIlwraith; 39.5% in Hinkley and Drysdale, and 43% in Fawkes *et al.*

Patients' ethnic status was an optional question and was fully completed in the PROM app pilot. Data collected showed that 80.5% of participants described themselves as White British, and 9.7% as White "other". A total of 6.8% of participants described themselves as belonging to non-White ethnic groups. In the 2012 study by Fawkes *et al*. 93.9% of participants described themselves as white so the pilot has included patients from more ethnically diverse backgrounds.

Employment status was discussed in previous data collection studies. In the PROM app pilot 67.7% of patients were employed full or part time compared with 76.1% in the study by Fawkes *et al.,* 2012. The duration of patients' symptoms has been described in a range of studies. In Pringle and Tyreman 64.5% of patients had experienced their symptoms for six weeks or less, and 23.9% for 11 weeks or more; 40.4% of patients in the study by McIlwraith reported symptoms of four weeks or less, and 6.3% for 4-6 months; and Fawkes *et al*. reported symptoms of 0-6 weeks in 51.1% of patients, 7-12weeks in 14.5% of patients, and 13 weeks or longer in 32.5% of patients. In contrast this study identified patients reported symptoms of 0-6 weeks in 38.9% of cases, 7-12 weeks in 9.7% of patients, and 13 weeks or more in 51.4%. This suggests that patients consulting osteopaths tend to have more chronic symptoms but this particular data may be skewed towards chronic patients as those patients with chronic and unchanging symptoms were being sought deliberately to participate in the test-retest strand of the study.

Across all studies low back pain was the most common complaint among patients 36% (Fawkes *et al.,* 2012), 52% (Burton, 1981), 68% (McIlwraith, 2003), and 49% (Hinkley and Drysdale, 1995). In the PROM app pilot the figure for low back pain was high (58.8%) but once again that may be influenced due to the fact that patients with low back pain were being recruited deliberately.

Service data were examined and indicated that access to osteopathic care is still timely with 22.2% of patients being offered an appointment the same day and 61.1% within three days of contacting a practice. In the study by Fawkes *et al*. 16.8% of patients were offered an appointment the same day and 71% within three days of contacting the practice.

Outcome data were collected in very different ways in this PROM pilot and in the study by Fawkes *et al.* In the latter, VAS data were collected by the clinician asking patients to mark a VAS scale, and by completion of a global improvement scale (Kemler *et al.,* 2003). A total of 74.3% of patients reported they were improved or much improved. Although the scale used in the PROM pilot is an alternative to the Kemler scale, and the patient population is slightly different, 81.3% of the patients in the responsiveness strand reported they were completely recovered or much improved; this was 55.6% in the test-retest strand, and 75% in the feasibility strand. These are heartening findings indicating that patients' responses are comparable when collected with the clinician or independently. However, the much smaller sizes of the data mean the findings should be treated with caution until larger data collection confirms or refutes these findings.

Satisfaction and experience data were collected in the PROM app pilot. Although these data have not been collected in previous data collection studies, there have been a small number of studies which have investigated patient experience and satisfaction with osteopathic care. Patient satisfaction has been shown to be high even when outcomes of care have been smaller than anticipated (Pincus *et al*., 2000). All studies including this one have highlighted high levels of satisfaction (Licciardone *et al.,* 2002; Licciardone *et al*., 2001; Fawkes, 2005; Strutt *et al.,* 2008). The reasons for these findings have been explored and patients have noted good communication,

clinician empathy, the opportunity to ask questions about their symptoms and their management, and competence has been highlighted. In their study of patient experience, Drysdale *et al*. identified that 88.7% of patients reported good experience of osteopathic care either within osteopathic educational institutions or in private practices (Drysdale *et al*., 2013). In this study 60% of patients reported very good experience when measured across all strands of the study with individual strands having higher ratings, and 62% of patients reported very good experience across all strands, although once again individual strands of the study have higher ratings.

Responsiveness

Values for the area under the curve (AUC) exist for studies involving the RMDQ. A range of values have been reported ranging from 0.64 (Maughan *et al.,* 2010) to 0.93 (Beurskens *et al.,* 1996). The values calculated for this study (0.732 at one week post treatment, and 0.788 at six weeks post treatment) fall in the middle of that range and are comparable to the value (0.76) identified by Davidson and Keating in 2002. Other values calculated for the AUC for the RMDQ include 0.82 (Coelho *et al*., 2008), 0.69 (Frost *et al*., 2008), 0.89 (Grotle *et al.,* 2004), 0.84 (Mannion *et al*., 2006) and 0.85 (Stratford *et al.,* 1994).

When addressing the values for the BQ, this study calculated the AUC for the BQ at six weeks post-treatment was 0.881 (95%CI 0.698 to 1.00). In earlier studies using the BQ, AUC values of 0.80 (95%CI 0.73 to 0.86), and 0.69 (no CI values available) were calculated for the BQ (Bolton and Hurst, 2011; Perillo and Bulbulian, 2003). This makes the one week and six week values from this study comparable.

AUC values for the VAS have been produced in many studies. While this study identified a value for the AUC for VAS as 0.670 (95%CI 0.412 to 0.929) at one week post treatment, and 0.972 (95%CI 0.892 to 1.00) at six weeks post treatment other values have been published. Recent studies examining low back pain patients have reported AUC values for VAS to be 0.76 (95%CI 0.68 to 0.83) and 0.73 (CI not available) (Janwantanakul *et al*., 2015; Parker *et al*., 2011).

Interpreting Change

Comparison of values for the MIC is more challenging since there have been fewer studies published where this calculation has been made. For the BQ more studies have been undertaken involving patients with neck pain and values for MIC of 5 (Perillo and Bulbulian, 2003), and 26 for acute patients and 18 for subacute (Newell and Bolton, 2010) have been published. The value for MIC calculated in this study was 12.5 which falls outside the subacute value suggested by Newell and Bolton, and the value of 5 suggested by Perillo and Bulbulian (Perillo and Bulbulian, 2003; Newell and Bolton, 2010). The scores were gathered from patients with symptoms lasting under six weeks so a MIC closer to 26 would have been expected. The baseline BQ sum scores in the responsiveness strand ranged from 0 to 63 (35% were 0-10; 18% were 11-20; 35% were 21-30; 3% were 31-40; 4% were 41-50; 1% were 51-60, and 1% were 61-70 , and there were no higher scores than 63). When addressing the value of the MIC for the BQ, it was calculated to be higher than the SDC (7.72) so we can have confidence that the PROM system is measuring "real change" outside of measurement error for the PROM.

When addressing the MIC for the RMDQ, a range of values have been calculated using different methodologies. These include 2.0 to 8.6 points, with 11-13 point published in studies for high baseline scores (Beurskens *et al*., 1995; Kopec *et al*., 1995; Stratford *et al*., 1996; 1995; Stratford *et al*., 1997; Johansson and Lindberg, 1998; Riddle *et al.,* 1998; Wiesinger *et al*., 1999; Roland and Fairbank, 2000; Stratford *et al*., 2000; Garratt *et al*., 2001; Nusbaum *et al*., 2001; Davidson and Keating 2002; Stratford and Binkley, 2002; Grotle *et al*., 2004; Jordan *et al*., 2006). These empirical studies were evaluated by Ostelo *et al*. using an intergrated approach and an absolute cut-off value of 5 points was recommended for a 30% improvement in MIC from baseline (Ostelo *et al*., 2008). In the responsiveness strand the change reported by patients was 27.59%, and a value for MIC of 4.5 was calculated. This is close to agreement with the recommendation from the expert panel described by Ostelo *et al.* considering the baseline scores in this study ranged from 1-20 with a mean of 6.2 reported (Ostelo *et al*., 2008). The value for the SDC was calculated to be 1.29 in this study suggesting that the change measured by the PROM in the app reflects "real change" *i.e*. beyond measurement error.

Values for MIC for the VAS were discussed also by the expert panel as described in Ostelo's paper (Ostelo *et al.,* 2008). The expert panel once again examined empirical data and identified for the VAS with a scoring range of 0-100, absolute MIC values of 2.0 to 29.0 points (Beurskens *et al.,* 1995; Hagg *et al*., 2003; Grotle *et al.,* 2004). No data concerning percentage improvement from baseline were available to complement the MIC values identified. On discussion a final cutoff value of 15 was recommended for the MIC using the VAS with a 30% improvement from baseline recommended also. In this study a mean change of 33.31 was identified between baseline and 6 week follow up, and a MIC value of 27.50 higher than the recommended cut off but just within the limits of the empirical studies identified in the Ostelo study (Ostelo *et al*., 2008). Participants in this strand of the study had symptoms of 0-6 weeks' duration with a mean pain score of 47.39, so a higher MIC score could be anticipated. The SDC for the VAS was calculated to be 7.56.

Reliability

Using COSMIN guidelines and recommendations, test-retest reliability was evaluated by calculating values for the intraclass correlation coefficient (ICC) and for Cronbach α. In this study, values calculated for the respective PROMs were

	BQ	RMDQ	VAS
ICC	0.520 (95%CI 0.045 to 0.767)	0.653 (95%CI 0.260 to 0.839)	0.599 (95%CI 0.013 to 0.838)
Cronbach α	0.791	0.844	0.858

Among studies of the BQ involving patients with low back pain, previous values for reliability have been identified. These include ICC of 0.95 and Cronbach α of 0.9 (Bolton and Breen, 1999); ICC of 0.96 and Cronbach α of 0.89 (Hartvigsen *et al*., 2005). The value for ICC in this study was lower than in previous studies which used hard copy questionnaires instead of an electronic version. The value of Cronbach α is regarded as good. When considering the values for the RMDQ, more studies are available which have calculated values for ICC and Cronbach α while evaluating test-retest reliability. Values for Cronbach α range from 0.38 (Frost *et al.,* 2008), 0.47 (Kopec *et al*., 1995), 0.49 (Davidson and Keating, 2002), 0.56 (Stratford *et al.,* 1994),

0.67 (Mannion *et al*., 2006), 0.72 (Beurskens *et al*., 1996), and 0.75 (Grotle *et al.,* 2004).

In studies calculating the ICC, values ranged from 0.76 (Patrick *et al.,* 1995), 0.88 (Johansson and Lindberg, 1998); 0.94 0.91 (Kopec, 2000), (Nusbaum *et al.,* 2001), 0.53 (Davidson and Keating, 2002); 0.89 (Underwood *et al.,* 2011), and 0.93 (Jacobs *et al.,* 2015). The value for ICC in this study is lower at 0.653. The only other study which has tested equivalence between paper and electronic versions of the RMDQ did not calculate values for ICC or Cronbach α so it is disappointing not to be able to compare the findings of this study to its nearest comparator (Bishop *et al*., 2010).
Studies evaluating the test-retest reliability of the VAS have identified good reliability. In their study comparing the test-retest reliability of the VAS when using a touch screen (iPad), Bird *et al*. calculated a value for the ICC of 0.90 (95%CI 0.82 to 0.95) (Bird *et al*., 2016). Their study involved 22 healthy older adults who completed the VAS on four occasions: two using the VAS on the iPad, and two occasions marking a paper version of VAS. The difficulty in comparison with this study is that the patients were not in pain, although whether measures of minimum important change in pain scales vary greatly between populations is disputed by some researchers (Kelly, 2001).

5.6.2 Strengths and limitations of this study

This is the first study which has evaluated the reliability and responsiveness of the BQ when used in an electronic format. Although the RMDQ has been compared for electronic and paper media, test-retest (TRT) reliability and responsiveness were not evaluated (Bishop *et al*., 2010). More recent work by Bird *et al*. has investigated the use of VAS in an electronic format and compared this with papers versions (Bird *et al.,* 2016). The use of the PROM app to collect independent outcome data is a new departure for the osteopathic profession. Earlier work which collected outcome data relied upon completion of a VAS by the clinician, or in some cases by the patient in the presence of the clinician which limited the robustness of the data (Fawkes *et al.,* 2012). The BQ has been used in a small number of osteopathic studies, and the findings of this study are comparable to this study even though the setting (an NHS

secondary care setting) was different to the private practice setting used in the pilot of the PROM app (Gurden *et al*., 2012).

The descriptive data gathered in the PROM pilot is useful for complementing earlier studies, and it has the advantage of being patient rather than clinician completed.
Although there are strengths to this study, there are limitations also. Although this is pilot data, there is a threshold number of datasets required for clinimetric evaluation and this was not reached for this study. The data must, therefore, be regarded with some caution. The recommendation by the COSMIN group is for a minimum of 30 data sets but a preference for 50 to be used for calculation of reliability statistics (Altman, 1991; Terwee *et al.,* 2011). Although recommendations for these calculations are not definitive, the recommendations are based on the considerable experience of the COSMIN group members and are a useful guide. As a result of not meeting this threshold, data will continue to be collected outside of the PhD to have a more robust basis on which to statistically evaluate ICC and Cronbach α values for the BQ, RMDQ, and VAS when used electronically. Within the sample of patients who contributed data for the TRT strand of the study issues were identified. Participating osteopaths were requested to invite patients with chronic and unchanging symptoms of non-specific low back pain to participate in the study. It has been apparent from the analysis that the patients did not have unchanging symptoms and this will have had an impact on the TRT data. At analysis, patients who had recorded "no change" on the transition question were included in analysis for the test-retest population. The "stable population" in the test-retest strand was used to try to recruit a suitable number of patients from osteopathic practice for analysis, and reduce the administrative burden on practices. In hindsight a more suitable method would have been to identify an osteopathic patient population with stable pain as described earlier, administered the questionnaire including the PROMs to patients one week prior to treatment, and administer the second (retest) questionnaire at their consultation prior to any treatment taking place (Bolton and Breen, 1999). In this way there would have been no intervention which could impact on patients' symptom pictures.

In the responsiveness strand of the study the same issue of small sample size was identified. In some cases baseline and one week data were collected, and in others,

baseline and six week data were collected. In too many cases complete datasets were absent. Recruitment of participants was generally slow throughout the study and this has impacted on the amount of data collected.

5.6.3 Future directions

Future studies will focus on building on the data collected already. A request to the charity BackCare has been met positively and recruitment of patients is continuing to contribute data for both TRT and responsiveness strands of the study. Larger sample numbers will allow re-calculation of clinimetric data which will be submitted for publication in peer-reviewed journals. Further development of the app will allow data to be collected from different body sites, and for different conditions. This will be discussed in more detail in Chapter 7.

5.6.4 Conclusions

The app performed well during the pilot from a functional perspective. Analysis of the data identified that questions were answered well and with good completeness suggesting the content is not regarded as burdensome by patients. The Roland Morris Disability Questionnaire was less frequently completed than the other PROMs and has been removed from the final version of the app.

6

Implementation of the PROMs app into day-to-day clinical practice

6.1 Introduction

In the field of healthcare new innovations are being introduced each year which could have the potential to enhance a patient's experience of that healthcare. While those new innovations may have been developed with great attention to detail, and in an evidence-based manner, if their implementation is poorly considered, they may not become part of day-to-day practice (Wensing, 2000). The adoption of new initiatives and procedures often takes place with a lack of completeness and with difficulty but there are some fundamental issues to be considered, and which are informed by the growing body of evidence from implementation science. Perhaps some of the uncertainty can be addressed by considering what is meant by implementation. It was described in 1997 by ZON (Zong Onderzoek Nederland), the Dutch national organisation for health research and healthcare innovation as "a planned process and systematic introduction of innovations and/or changes of proven value; the aim being that these are given a structural place in professional practice, in the functioning of organisations, or in the healthcare structure" (Hulscher *et al*., 2000; Grol *et al*., 2013).

In this chapter, I will review the theoretical bases behind implementation science, and how this insight has informed the implementation of the PROMs app (the development of which has been described in chapters 2,3,4 and 5) into osteopathic practice. The early findings of the implementation phase of this study will be described also, and the chapter will conclude with future directions for this initiative.

6.1.1 The challenges of implementation and change

Good patient care is the implicit aim of healthcare professionals, but in many instances good care may fail to materialise (Edmondson, 2004; Department of Health, 2013). There can be many reasons for this including failures of leadership, lack of competency, the discomforting challenge new research can bring to embedded practice and experience, lack of resources for care delivery, or failure to reflect on current standards of practice and outcomes of care (Berwick and Hackbarth, 2012). Behind each of these reasons lie further explanations, and it is in tackling these explanations that care might be improved.

I described in Chapter 4 that one challenge for healthcare professionals in the 21st Century is dealing with the sheer volume of new information produced each year. This information can, at times, be contradictory, and it becomes obsolete very quickly in the face of new scientific and social developments. One might reasonably question whether the time and money invested on the development of evidence to inform guidelines, and other underpinning initiatives to deliver good patient care is worth it when so much of what is produced fails to be implemented. This is not a new phenomenon. Although Ignaz Semmelweis demonstrated the importance of antiseptic practice in the 19th Century, campaigns still have to be undertaken today to encourage healthcare professionals to wash their hands when caring for patients despite their knowledge of the adverse consequences for those patients, and healthcare environments (Semmelweis, 1860; Bolon, 2011). While the literature on failure to implement optimal care is by no means scant, evidence that unnecessary, out-of-date, and ineffective care is frequently offered is abundant (Bodenheimer, 1999; McGlynn *et al.,* 2003; Asch *et al.,* 2006). The delivery of patient care which is timely, efficient, and patient-centred, requires organisation, reflection, and knowledge of best evidence. While patient-centred care requires the patient to be

involved in their care and the decision-making process, this change in focus concerning the role of the patient can represent a challenge to some clinicians who favour a more parentalist or overseeing approach (Coulter and Collins, 2011). Furthermore, although variations in care can occur between different professional and regional settings; they can be evident also within different professions (Engels *et al*., 2006). While healthcare professionals both nationally and internationally might agree that good healthcare delivery is:

- effective;
- safe;
- patient-centred;
- efficient;

the manner in which this should be achieved is widely debated by different professional and organisational stakeholders including patients, clinicians, researchers, advocacy groups, manager, ethicists, lawyers, politicians, insurers, and other third-party funders. New strategies are introduced to try and implement changes in good practice but in some instances it can be clear from existing evidence when implementation has failed (Peute *et al*., 2010; Haydon, 2013). Although it would be more straightforward if there was agreement on what was the most effective manner of implementation. Approaches to changing clinical practice can include, for example:

- A cognitive approach;
- A motivational approach;
- A marketing approach;
- Approaches reinforcing behaviour;
- A social interaction approach;
- A management approach;
- A control and compulsion approach (Wensing *et al*., 2000).

Each of these approaches is described more fully in Table 6.1.

Table 6.1. Approaches aimed at improving clinical care (Kitson *et al*., 1998; Wensing *et al*., 2000).

Theoretical approach	**Description of the underlying approach**	**Supporting resources required**
Cognitive approach	Decisions are made on a rational basis evaluating evidence and professional experience.	Summaries of evidence, and clinical guidelines where available.
Motivational approach	The importance of self-directed motivation underpins this approach driven by the desire to reach and deliver optimal standards of care.	Resources which underline the importance of clinical experience and how this can be translated to care *e.g.* problem-based learning, and changes driven from the bottom-up involving patients, and clinicians.
Marketing approach	Engaging stakeholders to identify an appealing strategy or proposal for change is key to this approach.	Involving different stakeholders to create the most effective messages, and then accessing a range of communication channels to transfer those messages *e.g.* clinical networks, and "champions".
Reinforcing behaviour	This approach, based on the assumption that behaviour is influenced by external factors, is underpinned by learning theory, and economic principles of reinforcement and conditioning.	Strategies which underpin this approach include different feedback mechanisms including peer-to-peer feedback, and individual feedback compared to a group norm/standard. Incentives or sanctions can be involved in this approach *e.g.* performance-related

		pay.
Social interaction approach	The influence of others is an implicit part of this approach. This can include respected colleagues, local or national opinion-leaders, or others in a respected authority position. Learning and change is believed to be facilitated by contact with colleagues who demonstrate high standards of behaviour in the delivery of care.	The role of the team is a key resource in this approach providing feedback, opportunities for informed discussion, and peer-feedback. Outreach visits by respected opinion leaders are also regarded as helpful.
Management approach	This approach is based on a more directive approach with tiers of management identifying and changing processes and structures identified as underpinning poor performance.	Continual monitoring of changes introduced is the main resource underpinning this approach.
Control and compulsion approach	This approach is underpinned more heavily with sanctions when poor performance is identified.	Financial sanctions to individuals and organisations in the form of salary penalties, contractual, and licensing penalties are more commonly used resources.

Notwithstanding the different approaches listed in Table 6.1, two contrasting approaches have been broadly identified to the implementation of knowledge or process changes in healthcare (Kitson *et al*., 1998; Van Woerkom, 1998; Hulscher *et al*., 2000; Wensing *et al*., 2000) as shown in Figure 6.1.

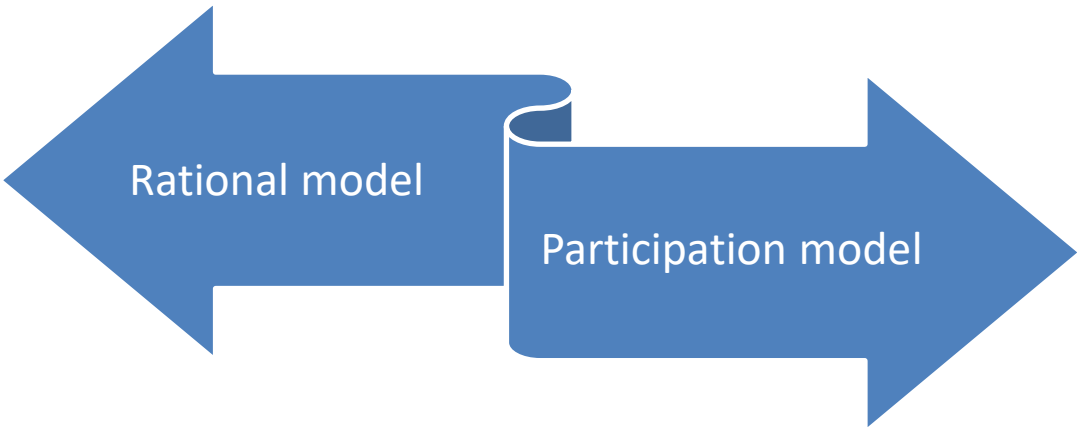


Figure 6.1. Broad models of implementation of knowledge or process change in healthcare (Kitson et al, 1998; Van Woerkom, 1998; Hulscher *et al*., 2000; Wensing *et al.*, 2000).

Each model has its own distinctive features, and the features of the rational and participation model are developed in Table 6.2.

When considering implementing changes to practice, elements of both approaches tend to be used within the existing structures of a particular setting. This view is reinforced by many regulatory authorities who advocate that "optimisation of patient care may be seen as a two-way flow between practice and science" (Health Council of the Netherlands, 2000; Grol *et al*., 2013). It has been with this approach in mind that the PROM app content development was based on the use of qualitative input from patients and clinicians, and best available evidence for measurement tools. When considering different implementation strategies and approaches, it becomes clear that there is no such thing as one clear and effective strategy for all settings, despite the desire from many healthcare stakeholders for a "one-size-fits-all" strategy. It has been proposed that different settings and populations require different methods of implementation (Grol and Grimshaw, 2003; Grol and Wensing, 2004). In the next

section, I will explore the different implementation strategies recorded in the literature in greater detail.

Table 6.2. Features of the rational and participation model approaches to implementation (Van Woerkom, 1998; Grol *et al*., 2013).

Rational model	**Participation model**
A clear starting point is evident;	The starting point is less clearly defined;
The starting point is based on the availability of new evidence, insights, or procedures agreed to be of value to the population in question;	Implementation occurs in a step-by-step manner, but the innovation is not always introduced in an evidence-based process;
Steering takes place externally, and, in the main, in a top-down manner;	Changes occur in overlapping stages whereby the development, testing, and introduction of the innovation intertwine;
Increased adoption of new innovations by target group members;	Inter-professional discourse by staff using the new innovation on a daily basis determines whether or not its implementation succeeds;
Focus is on the most prevalent needs with less attention to diverse needs in the population;	There is a lack of awareness about the need for change to be introduced;
Little use is made of existing experience and niche knowledge present within the target population	Little attention is paid to existing structures and how they can affect implementation;
Implementation driven by the supply of technology;	Implementation is driven by the need for technology;
Target groups are frequently positive about new innovations.	Target groups are generally neutral about new innovations.

6.1.2 Theories underpinning implementation and behaviour change

While the implementation of change in clinical practice is largely intended to improve care and performance, in many cases this is not achieved easily. A number of factors need to be considered when changes are proposed including:

- The innovation – is it a credible proposal for the target population, and has it been developed or been based upon high quality evidence?
- The target population – what are the characteristics of the population in terms of their knowledge of the innovation? What is the necessity for change, and having the skills to adapt to that change?
- The professional culture – what are the existing norms for the profession, are they consistent across the profession or are there significant intra-professional differences which need to be understood and accommodated?
- The patients – what are their attitudes to innovation in practice: will this meet expectations or be burdensome? What are their knowledge and skills which will support or hinder compliance?
- The practice setting – what will the attitudes of colleagues be like to the new innovation? Will they regard it as useful or useless, and will it be perceived as a threat to their autonomy or ability?
- The economic context – could there be financial pressures through introducing the innovation, or as a consequence of the innovation itself?;The regulatory context – the new innovation may be regarded as a top-down directive from a regulatory authority which may mean that the innovation is introduced albeit grudgingly;
- The professional organisational context – once again this might be perceived as a top-down directive about an innovation for which merit has not been proven, or it may be regarded as an attempt to support and promote an aspect of professional care;
- The dissemination strategies available to the professional group – this can vary between professions depending on their normal modes of communication, opportunities to engage in Continuous Professional Development (CPD) activities, and levels of engagement in new technologies which can support dissemination (Grol, 1992; Wensing *et al*., 2010; Grol *et al*., 2013).

Implementation theories try to explain the factors that are important and contribute to the success of the implementation of an innovation. A range of different theories exist and depend on the setting, political and economic context, population, or innovation itself. Rossi *et al*., 1999 described theories as being either due to impact or process. The elements of impact and process theories as applied to healthcare professionals are summarised in Figure 6.2 (Grol, 1997; Wensing *et al*., 2010). These theories focus upon the manner in which healthcare professionals make decisions or consider choices (cognitive theories), the attitudinal factors behind manifesting a change in practice (educational theory), and the strategies for realising an improvement in their performance (motivational theory). It can be argued also that social and organisational context can be important factors in implementing changes but these have resonance only when they are perceived to be important by individuals (Grol, 1997; Wensing *et al*., 2010). Each of these theories will be considered in turn.

Cognitive theories

Rational processes of thinking and decision-making need to be examined when attempting to introduce change for healthcare professionals (Breuhaut *et al*., 2007). Categorised under the heading of rational decision-making theories, these theories assume an analytical approach whereby healthcare professionals evaluate the potential advantages and disadvantages of a particular innovation and its alternatives on their ability to provide high standards of patient care. Clear and evidence-based information is required to support this decision-making process *e.g* clinical guidelines, or primary research. This will be weighed against the perceived effects of not implementing the change, and the effect that could have on patient care. Resource demands in terms of time, individual capability, and workload must be considered also.

While rational decision-making identifies what can be seen to be important factors in the implementation process, other cognitive theories focus more extensively on how decisions are made *e.g.* using a cognitive-psychological approach. This approach states that clinicians may not in fact act rationally but their actions are based upon their previous clinical experiences within a particular context. The use of "illness

scripts" has been identified where clinicians have been shown to organise their knowledge in a particular way while accessing previous experiences as an implicit part of the decision-making process (Hobus, 1994; Norman *et al*., 1996).

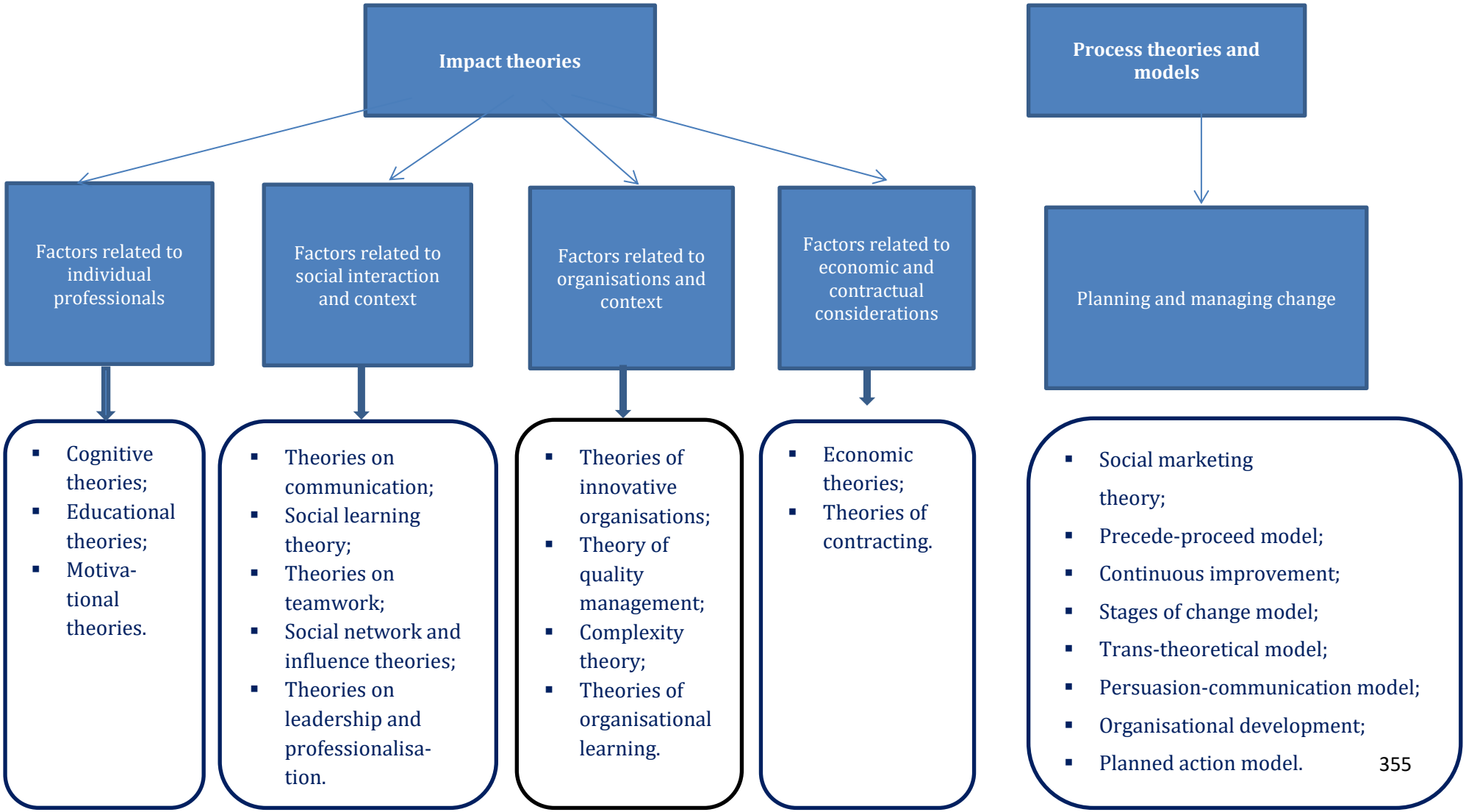
The more experienced the clinician, the greater number of illness-scripts he/she could access thereby allowing quicker diagnoses to be achieved (Schmidt, 1984). Festinger initially identified this behaviour, and identified that some clinicians will cling to behaviour patterns which demonstrate consistency in thinking and acting even though the basis for this is flawed in some situations (Festinger, 1954). It is important to introduce evidence to support new interventions which will challenge this type of unproductive behaviour offering a contemporary rationale for change. In the implementation of the PROM app, the cognitive approach would be based upon the need to identify changes which could impact upon osteopaths' personal practice, and practice settings, but also potentially meet the challenges posed by regulators and third-party funders (including healthcare commissioners and insurers) for example.

Educational theories

Among the wide-ranging and numerous educational theories that exist, the underlying focus is on the motivation within the individual to learn new skills, reflect upon those skills, and endeavour to change how they practise where appropriate. Within the osteopathic profession, as with many others, there exists a range of educational skills and attainment, and exposure to different educational methods (Overton *et al*., 2009; Kursukar *et al*., 2012).

Within this same group there will also be a broad range of preferred learning styles (Stanley *et al*., 1993; Hawthorn *et al*., 2009; Taylor and Hamdy, 2013). Adult learning theories base their rationale on the idea that most adults will be more inclined to learn and change their practice when they are confronted with an experience or clinical problem to which they can translate new knowledge (Armson *et al*., 2007; Andersen *et al*., 2008; MacCarthy *et al*., 2012). Abstract information without any defined focus is less likely to be effective (White *et al*., 2004).

Figure 6.2. A summary of the elements of process and impact theories (Grol, 1997; Wensing *et al.*, 2010).



Educational theorists, Lewis and Bolden, identified four discrete personal learning styles:

- Activists – who will engage with new innovations very quickly, but also sometimes abandon them equally quickly;
- Reflective professionals – who will carefully consider all elements of a new innovation before engaging with it;
- Theoretical learners – who require rigorous information and analysis of the *status quo* to understand the rationale for introducing a particular innovation;
- Pragmatists – who prefer to act on an innovation if they have had experience of it or something closely aligned to it (Lewis and Bolden, 1989).

The individual learning styles represent important considerations for implementation of change. It may be necessary to tailor an implementation strategy for each learning style, or recognise that there may be elements of all learning styles within individuals although one might supercede others. In order to build on interventions supporting both educational and cognitive therapies, it is important not to underestimate the importance of motivation.

Motivational theories

Motivation is defined as "a motivating force or incentive" (Chambers, 1989). Motivation theories focus on many of the elements which can be encouraging or discouraging when a new intervention is being discussed. Such elements include individual and institutional attitudes to change (Bandura, 1986; Maibach and Murphy, 1995; Caldwell *et al*., 2008; Campbell, 2008; Mansouri *et al*., 2009), perceptions of change generally and their benefits or disadvantages (Kortteisto *et al*., 2010), and intention toward a particular action (Fishbein and Ajzen, 1975; Ajzen, 1988; Kok *et al*., 1991; Godin *et al*., 2008; Sassen *et al*., 2011). One of the most widely-recognised theories for changing practise is the Theory of Planned Behaviour (TPB) (Ajzen, 1980). TPB aims to predict an individual's intention to engage in a particular behaviour at a specific time and place: it was intended to explain all behaviours over which people have the ability to exert self-control. The key component to this theory is those behavioural intentions which are

influenced by how likely the behaviour will have the expected outcome, and includes also the subjective evaluation of the risks and benefits of that outcome.

Although TPB is widely recognised and used, it does not take into account whether the healthcare professionals concerned have had the opportunities and resources, *e.g.* training, to be able to implement a particular change (Knowles *et al*., 2015). Other considerations include the effects of the practice environment, financial constraints, and previous experience, either successful or unsuccessful (Watkins *et al*., 2015). These are important factors and the failure to take them into account undermines the theory by failing to account for factors which are unrelated to a person's intention to change. Although other theoretical models have been constructed, their integration with the TPB gives rise to the most commonly used model in implementing behaviour change (Taylor *et al*., 2006). Many models, however, fail to recognise that individuals can change over time, especially if they begin to see new innovations being successfully implemented and producing useful output (Walker *et al*., 2001; Francis *et al*., 2009). Mindful of the elements of the theories described, an implementation strategy has been produced for the introduction of the PROM app into osteopathic practice. This is described in Section 6.2.

6.2 Methods - Delivering the implementation process

While much can be learned from trials of behavioural change for new innovations, the challenge is to disentangle the effects of that intervention from other contextual factors that might also influence a particular behaviour change (Grimshaw *et al*., 1995). When attempting to introduce the use of the PROMs app into mainstream clinical practice there were essentially two issues to consider: the change in behaviour for clinicians in implementing its use into day-to-day patient care, and the change for patients in completing questionnaires periodically during their treatment interval. Reviews of evidence have attempted to identify the key ingredients for successful implementation but there is no magic formula as the interventions are context and population-specific. Different studies have addressed this conundrum. Bero *et al.* undertook an early review of

implantation studies and found that some common themes were identifiable with varying levels of effectiveness (Bero *et al.,* 1998) (Figure 6.3).

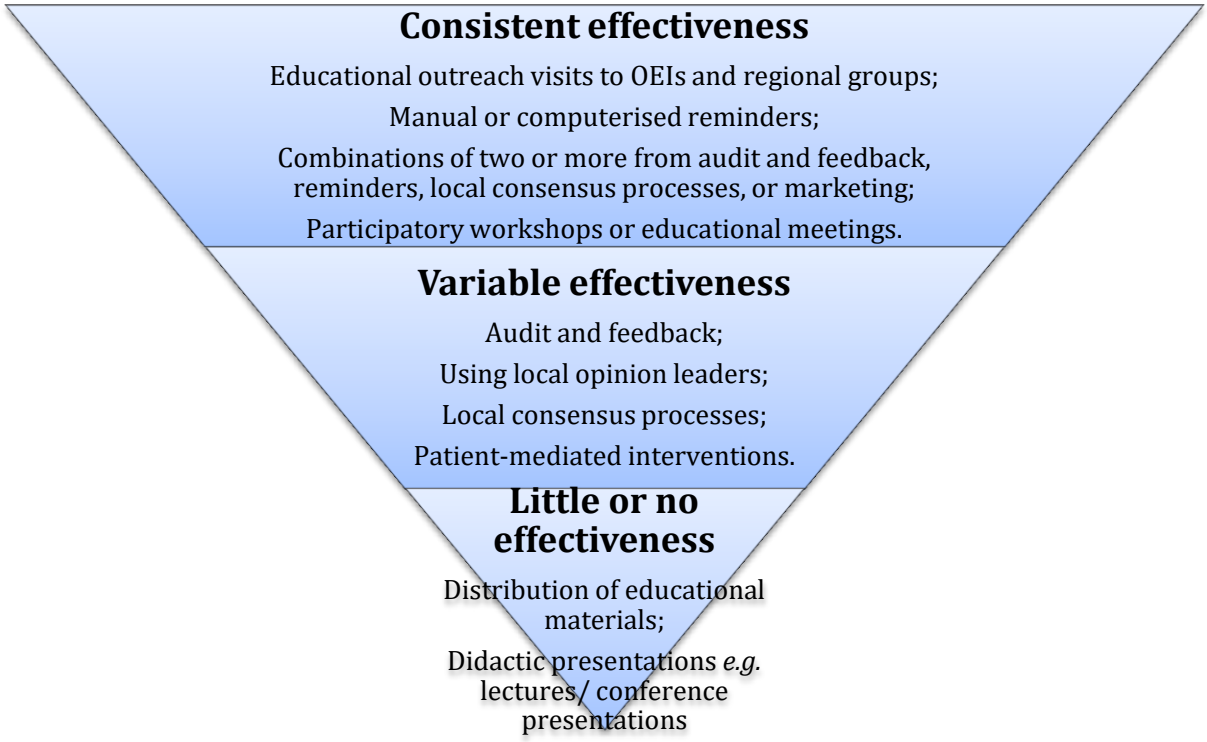


Figure 6.3 Interventions to support behaviour change and effective implementation in descending order of effectiveness.

While the review by Bero *et al*., 1995, and additional work by Oxman *et al.,* 1995 have identified the actual implementation strategies, few studies have actually compared the relative effectiveness between different types of intervention. An earlier review in the Effective Healthcare Bulletin examined 22 studies comparing different implementation measures but the generalisability of the findings, to different healthcare settings, levels of educational delivery, and populations, was questioned due to the heterogeneity of these studies (Effective Health Care, 1994). While there are considerable numbers of models for changing clinical practice, two broad models have been identified which underpin theories for implementing change with respect to knowledge, and processes to apply that knowledge *i.e.* the rational and participation approach. Within these broad models, a range of theories and strategies have been suggested to support implementation of the PROMs data collection facility into practice.

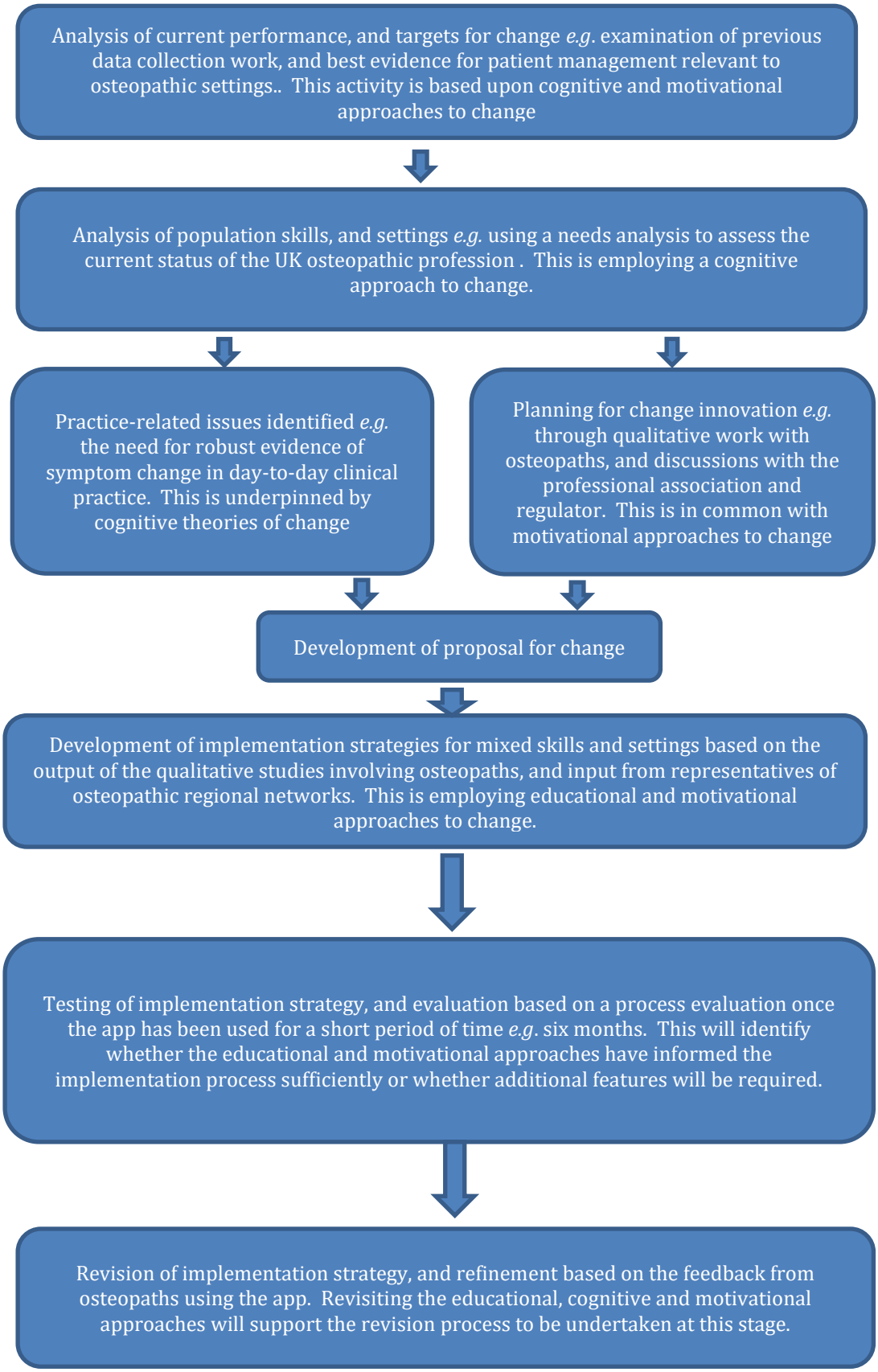


Figure 6.4 The Grol and Wensing Implementation of Change Model (Grol *et* 2013)

Creation of an implementation strategy

The strategy I have used has been focused upon the factors affecting individual professionals. These factors include cognition, education, and motivation. Breuhaut *et al*. identified that clear and evidence-based information is required to support rational decision-making in healthcare professionals (Breuhaut *et al*. 2007). The PROM data collection system was developed based on underpinning research and this has been made explicit to osteopaths through a range of communications, and to volunteers enquiring about using the app during its pilot stage and thereafter. The manner in which the PROM data collection system could have an impact on osteopaths' individual and collective practise must be communicated also through a variety of media to suit individual learning styles. This approach is supported by Lewis and Bolden, 1989 who described four distinct learning styles.

The final application of theory to practice is identifying motivational approaches to resonate with different osteopaths who may have very different reasons for being interested in PROM data collection *e.g.* reflection on practise, business promotion, or approaching new health provision sectors. Each of these competing factors were evaluated to develop an implementation strategy based on Grol and Wensing's Model (Grol *et al*., 2013) shown in Figure 6.4.

Grol and Wensing identify a series of key stages to implementation of an intervention into practice (Grol *et al*., 2013). This approach was used while also applying the theoretical approaches identified in section 6.1.2. When applying this model, one of my initial actions in creating an implementation strategy was an analysis of practice-related issues. Some of the key stakeholders in osteopathic practice, current and potential, relevant to this process are shown in Figure 6.5. Other practice-related issues concerning the implementation of a PROMs app were identified in the qualitative study involving osteopaths described in Chapter 3. These issues included:

- Time;
- Availability of information technology;

- Relevance of PROMs to osteopathic practice;
- Disruption to the consultation flow;
- Confidentiality;
- Ability to interpret the findings;
- Concordance or discordance with self-assessment of treatment capabilities;
- Potential patient burden;
- Commercial disadvantage (if patients disliked the idea of using the app);
- Expectation to conform (from patients and professional/regulatory bodies);
- Concerns about use of data;
- Concerns about data sharing.

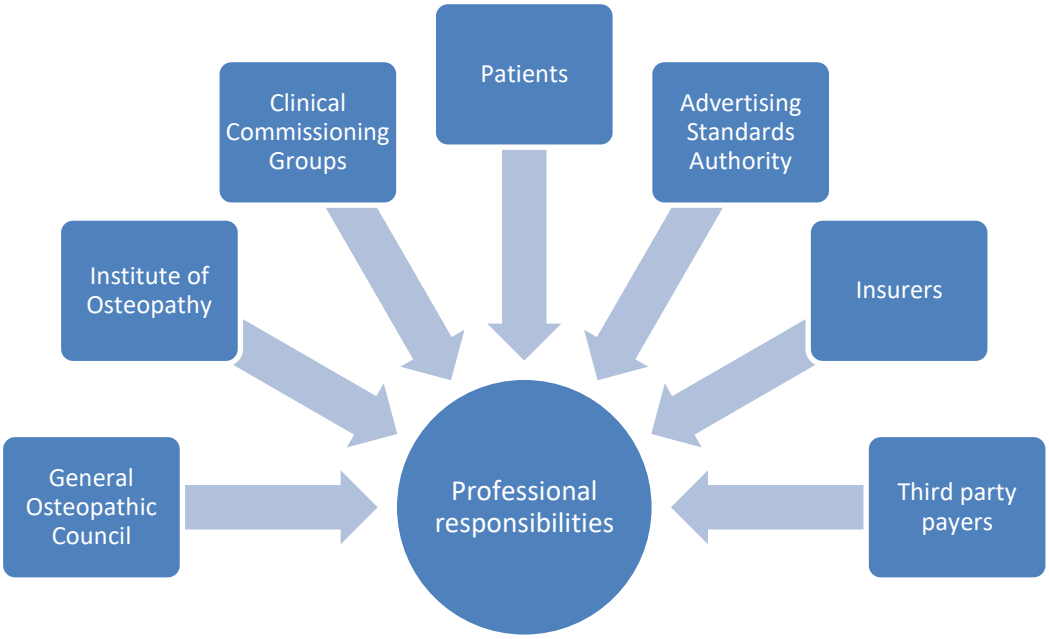


Figure 6.5. Key stakeholders in osteopathic practice

While acknowledging these issues, it was also important to consider the profession as a whole and the environments within which clinicians practise. All osteopaths must be registered with the General Osteopathic Council to be able to use the courtesy title "osteopath". A database of all registrants is held by the General Osteopathic Council (GOsC). It was not possible to examine the actual GOsC database due to the complex manner in which the database is structured (Tucker, 2015 –

personal communication), therefore the analysis of the environment in which osteopaths practise is drawn from mixed sources. According to the GOsC website

"*There are more than 5,000 osteopaths registered with the General Osteopathic Council, which includes some who practise abroad. The profession attracts almost equal numbers of male and female practitioners, and some have already qualified in another healthcare practice such as medicine, nursing or physiotherapy The majority of UK osteopaths (86%) practise in England, with 3.2% in Scotland, 2.4% in Wales, 0.4% in Northern Ireland and 8.4% working overseas. Most osteopaths are self-employed and work in the private sector, although some are working in multi-disciplinary environments within the NHS and in occupational healthcare in public bodies and private companies*" (GOsC, 2016).

The GOsC does not hold current data on the practice environment of the remaining osteopaths not included in those figures. The current GOsC database identifies that a small number of registrants are non-practising (n=151), the reasons for this are given in Table 6.3 (GOsC Registration report, 2016).

Table 6.3 Number of GOsC registrants non-practising with reasons

Reason for non-practising status	Number
Maternity/paternity	63
Ill health	19
Sabbatical	10
Travelling	7
Other*	52

Other includes: studying; not being able to find work; relocation of home/work premises; circumstances around the loss of a spouse/parent/child; acting as a carer; research; and pursuing other careers.

In a census undertaken by the Institute of Osteopathy in 2014, current data were recorded by the participating osteopaths (12% of the profession) shown in Table 6.4.

Table 6.4 Working environment for UK osteopaths participating in iO Census, 2014

Number of osteopaths	Number of practices where osteopaths work
59%	1 practice
28%	2 practices
13%	3 or more practices

This data broadly concurs with findings from survey work undertaken by the GOsC for their 2006/7 pilot practice survey (Figure 6.6), and by KPMG in 2011 (Figure 6.7) where 208 osteopaths participated.

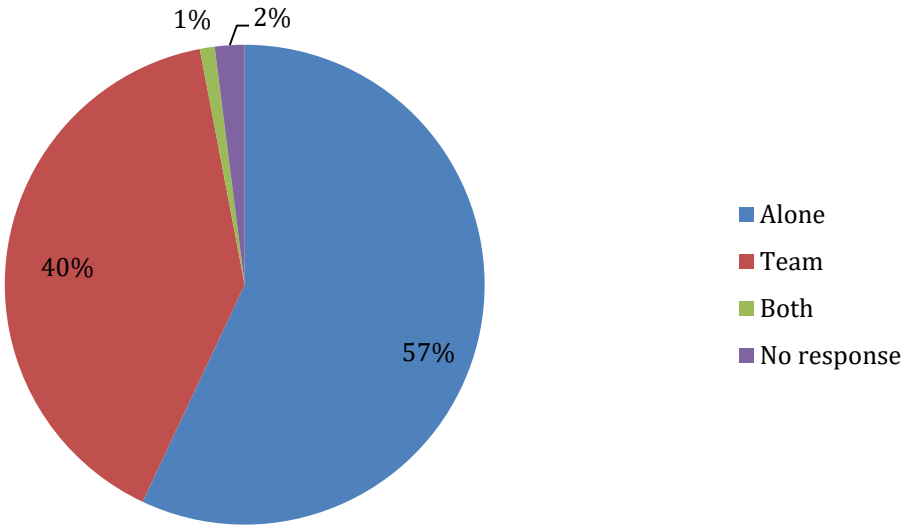


Figure 6.6 Working practices for osteopaths (2006/7 GOsC pilot practice

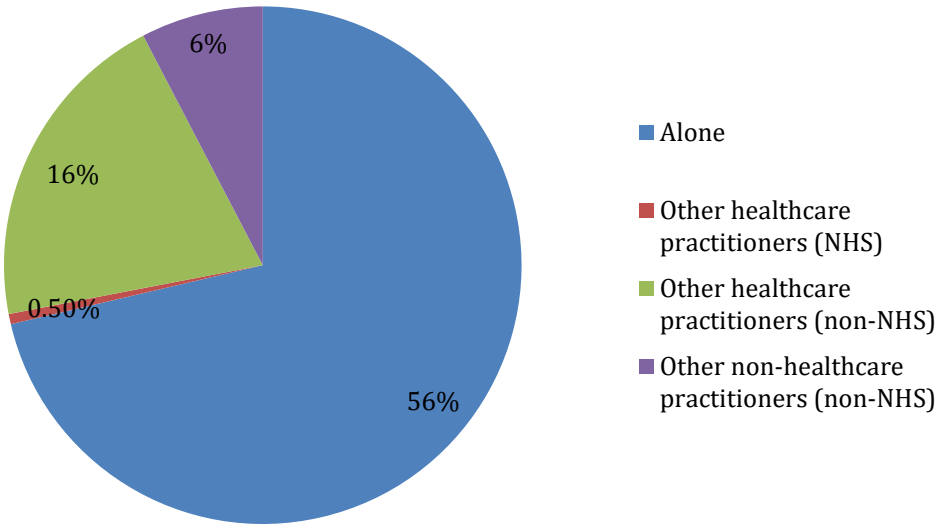


Figure 6.7 Working practices for osteopaths: where they spent ≥ 90% of their time

The survey work undertaken by KPMG in 2011, provided some information about working environment as shown in Table 6.5.

Table 6.5 Work setting for osteopaths in KPMG survey sample.

Practice setting	Percentage of sample (number of respondents in brackets)
Own home - room set aside as a clinic	12% (24)
Own home - room for clinical & domestic use	1% (3)
Own sole practice	22% (46)
A group practice	31% (65)
A surgery*	7% (15)

**this information does not distinguish whether this is working as part of the NHS, or whether this represents use of NHS premises on a rental basis.*

Work settings provide useful information about how a new innovation might fit into existing practice. Consideration needs to be given to whether there is a separate space at a practice to use an available desktop computer, laptop, or tablet computer to complete the PROM app should the patient wish to do so. In some practices there is no Internet connection at all. If the practice is part of an osteopath's home they

may also be unwilling to have people arriving or leaving treatment later than planned if this impacts on their family time or environment. When considering technology-based innovations, the issue of rural or urban practice settings is important also. This relates to availability of internet and speed of internet connections, mast coverage for phone signals, and population density in the surrounding area. The data collected by KPMG provided some data as shown in Table 6.6 (KPMG, 2011).

Table 6.6. Practice settings for osteopaths for 90% or more of their time.

Practice settings	Percentage of sample (number of respondents in brackets)
City-based areas	24.5% (51)
Town-based areas	37% (77)
Village-based/rural areas	10.6% (22)

Issues raised during the focus groups and individual interviews with clinicians (Chapter 3) identified the skills needed to be able to deal with PROM data, and previous experiences for some osteopaths of collecting practice-based data and not knowing quite what to do with it. Osteopaths are increasingly being expected to learn new skills and engage in new processes that they may be reluctant to engage with learning data analysis skills. Consideration was given to all of the factors mentioned as the implementation strategy was created as shown in Table 6.7. Issues raised by patients and clinicians were considered when creating an implementation strategy for the PROMs app, and these are shown in Figure 6.8.

Table 6. 7 Strategic considerations for PROM app implementation strategy		
Issues raised	Patients	Clinicians
Time to complete the app	The app contains a small selection of basic questions which can be completed within 5-10 minutes	The app is concerned with basic data collection avoiding excessive questions which could be perceived by patients to be burdensome.
Availability of information technology	Patients can choose to complete the app or not. They can do this using either a laptop, desktop, or iPad to access the web app. Alternatively, patients can download the mobile app onto an Android or iOS smartphone, or tablet computer.	Practices can choose whether to offer the option of using a practice computer or tablet.
Relevance of PROMs to osteopathic practice;	The PROMs included in the app were based on discussion with patients to consider relevance of content, format, and length.	The PROMs included in the app were based on clinimetric evaluation, and discussion with patients.
Disruption to the consultation flow;	The app can be completed at the practice or at home. It can be discussed during the consultation but not completed during it.	Interest in completion of the app can be raised by the clinician, or practice receptionist where applicable. Information is provided to patients to allow them to complete it in the environment of their choice.
Confidentiality;	No identifiable data will be shared. All data are encrypted via a Secure Sockets Layer (SSL), and never stored in an unencrypted form.	Aggregate data only will be provided for clinicians.
Concordance or discordance with self-assessment of treatment capabilities;	N/A	Patient feedback may highlight a practise-related issue which could be addressed by CPD.
Ability to interpret the findings;	N/A	Data will be analysed for clinicians and provided for them in summary form.

Potential patient burden;	The content of the app has been designed to be short allowing quick completion. The follow-up intervals have been selected also to avoid excessive demands on patient time (*i.e.* one week, and six weeks post treatment).	The choice of PROM in the app has been designed with patient input, and follow up time has been developed to avoid excessive demands on patient time.
Commercial disadvantage (if patients dislike the idea of using the app);	N/A	Completion of the app is entirely voluntary.
Expectation to conform (from patients and professional/regulatory bodies);	Patients indicated they were happy to give feedback when asked.	PROMs are being increasingly required from various stakeholders, and their use could contribute to opportunities for osteopaths.
Concerns about use of data;	Data will not be sold to commercial organisations or any other third parties.	Raw data will not be available to anyone except the research team involved.
Concerns about data sharing.	Individual patient data will not be shared with clinicians, it will be presented in summary form only.	Summary data will be available to individual clinicians, and their data will not be shared with other clinicians e.g. in the form of league tables *etc*.

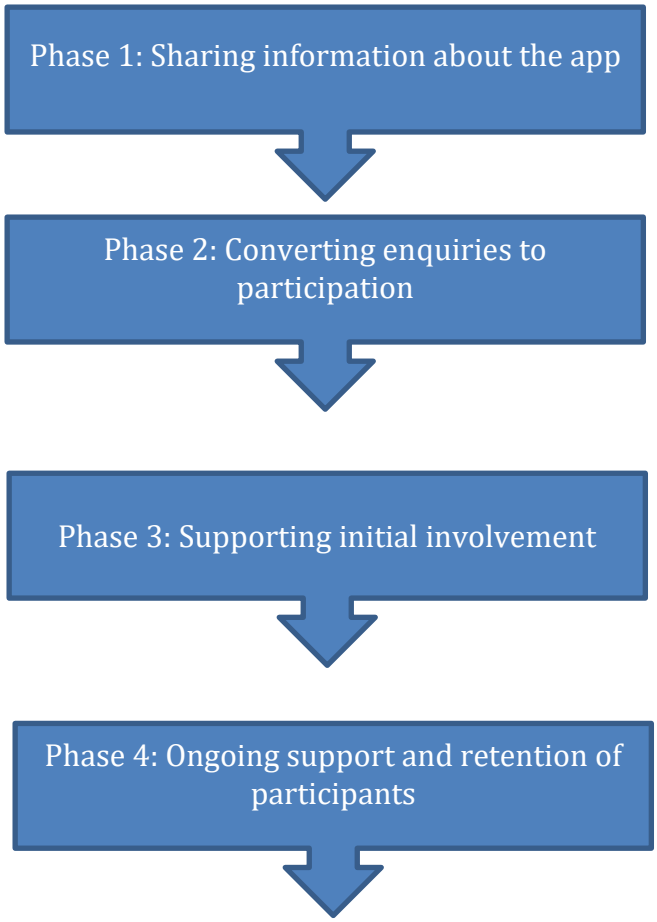


Figure 6.8 The strategy and relevant actions based on clinician and patient feedback.

Implementation activities

Mindful of the inconclusive nature of the literature concerning the value of multi-faceted or single interventions, a programme of activities was arranged. These are described in Phase 1 in Figure 6.8. In Phase 1, information sharing about the PROM app was included as part of normal outreach visits being undertaken by staff from the National Council for Osteopathic Research (NCOR) to keep the profession informed of its activities, published research relevant to practice, and sources of help and information where required. Articles in *The Osteopath* and *Osteopathy Today*, outreach visits and social media (Twitter and Facebook) were employed to give regular updates to the profession about how the app worked, the type of data collected by the app, and the potential benefits that could be obtained from using the app. The distribution of these activities is shown in Figure 6.9.

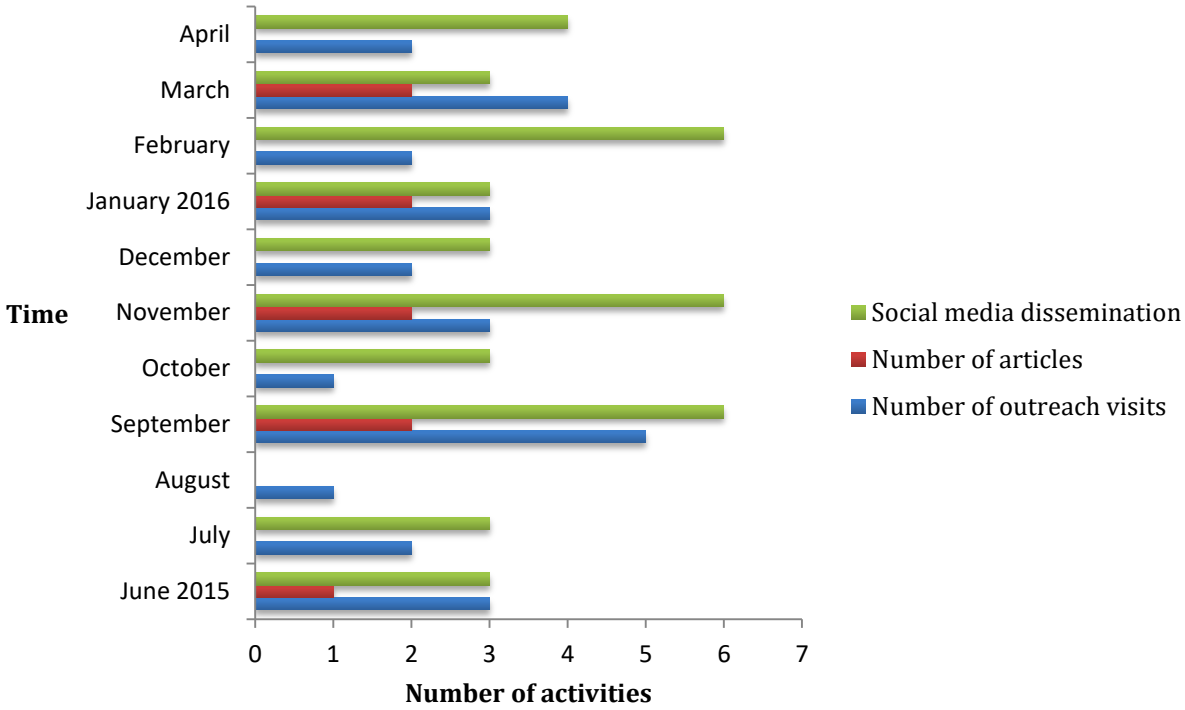


Figure 6.9. Activities supporting implementation of a PROM app into osteopathic practice

Information was included also on the NCOR website for osteopaths who wanted to further investigate using the PROM app without making personal contact. This information included more details about the developmental process of the app, a copy of Frequently Asked Questions for both patients and osteopaths, and other information about using the app in practice. In Phase 2, the aims and potential benefits of using the app for patients and clinicians were carefully considered ready to answer enquiries from interested osteopaths. This included the type of benefits for individual clinicians, how their data could be used as part of a pan-professional data set, and how their data could inform the development of their own practice from a business perspective, and practise from a clinical perspective through highlighting CPD opportunities.

In Phase 3, the focus was on the development of resources for osteopaths who volunteered to use the PROM app in their practices. The aim of creating the resources was to make the process as streamlined as possible thereby reducing potential administrative burden on clinicians or their practice staff. These resources

were created in response to the issues and strategies outlined in Table 6.7. The content of a resource pack is shown in Figure 6.10. In addition osteopaths received a poster to display in their practices describing the project, and containing a Quick Response (QR) code for ease of downloading the app. A QR code is a type of matrix barcode or machine readable optical label that contains information about the item to which it is attached (Gottesman and Baum, 2013; Lin *et al*., 2014). A QR code can be read by an imaging device *e.g.* a Smartphone or camera, and processed using Reed-Solomon error correction processes to extract data from the patterns embedded within the image (Riley and Richardson, 1998). A copy of the poster is shown in Figure 6.11.

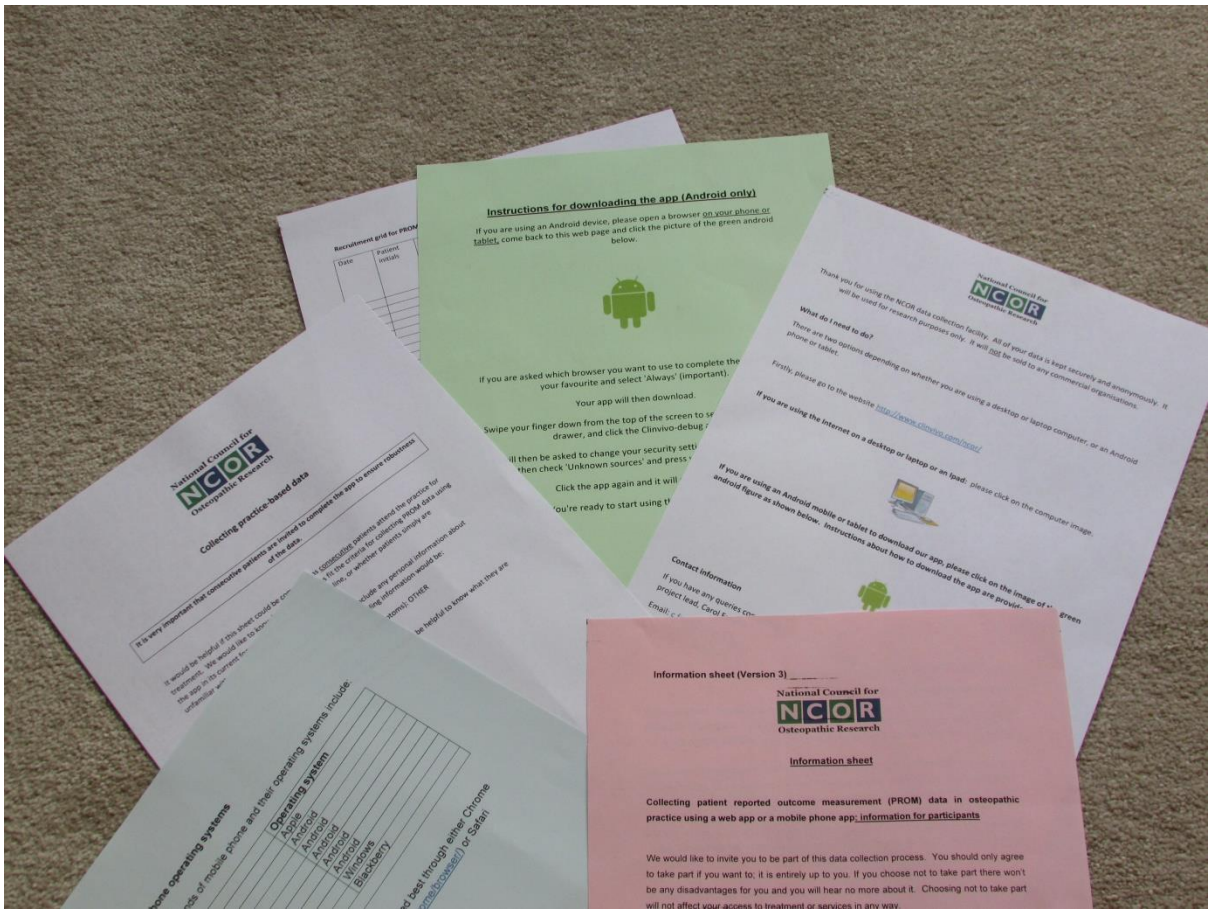


Figure 6.10. Content of resource pack provided for osteopaths using the PROM app.

Finally, Phase 4 focusses on the retention of volunteer osteopaths in using the PROM app. The ongoing feedback to osteopaths was concerned with data returns from

patients. The timing was agreed with individual osteopaths to ensure that enough feedback was provided to keep osteopaths informed about the number of patients submitting data and completing follow up questionnaires without anyone feeling that such feedback was being oppressive or intrusive. Some informal discussions have been held with osteopaths who have wanted to find out how to invite patients to take part in the process, and what pattern it should follow. The qualitative work involving members of the chiropractic profession helped immensely with this process because this gave me the opportunity to interview individuals who had used a similar system in their practices. The advice given was:

- Try to get the practice receptionists to do this (if that is relevant to the practice setting);
- Provide them with a "spiel" that has been tried and tested in clinics with experience of using electronic data capture;
- Ask them to enquire how patients have got on with the system, or whether they have received their follow up emails to remind patients to complete the questionnaires.

In the absence of a practice receptionist osteopaths have been advised to explain to patients that we are using a new system to collect data about the effects of treatment, and ascertain if they are willing to participate. If patients agree they are told about completing a short questionnaire on three separate occasions. The process for the post-pilot data collection is shown in Figure 6.12, and is available on the NCOR website (http://www.ncor.org.uk/practitioners/patient-reported-outcomes/prom-app-collecting-prom-data-in-practice/). If patients agree to participate in the data collection process, they are provided with a participant information sheet (PIS) and a participant code: the code is unique to the osteopath and the patient based upon the osteopath's regulation number: 399/002 was the second code issued by me. Further information about the data collection process, and importantly about the use of data, is available to patients on the NCOR website http://www.ncor.org.uk/patients/prom-app/. Patients have access also to an enlarged card providing information about how to access the app, information about

how to download an app, and information about the different operating systems for Smartphones (Appendix 6.1).

Figure 6.11 Poster to advertise the use of the PROM app to patients.

Although the implementation process has only begun recently, some findings are available, and these are described in Section 6.3.

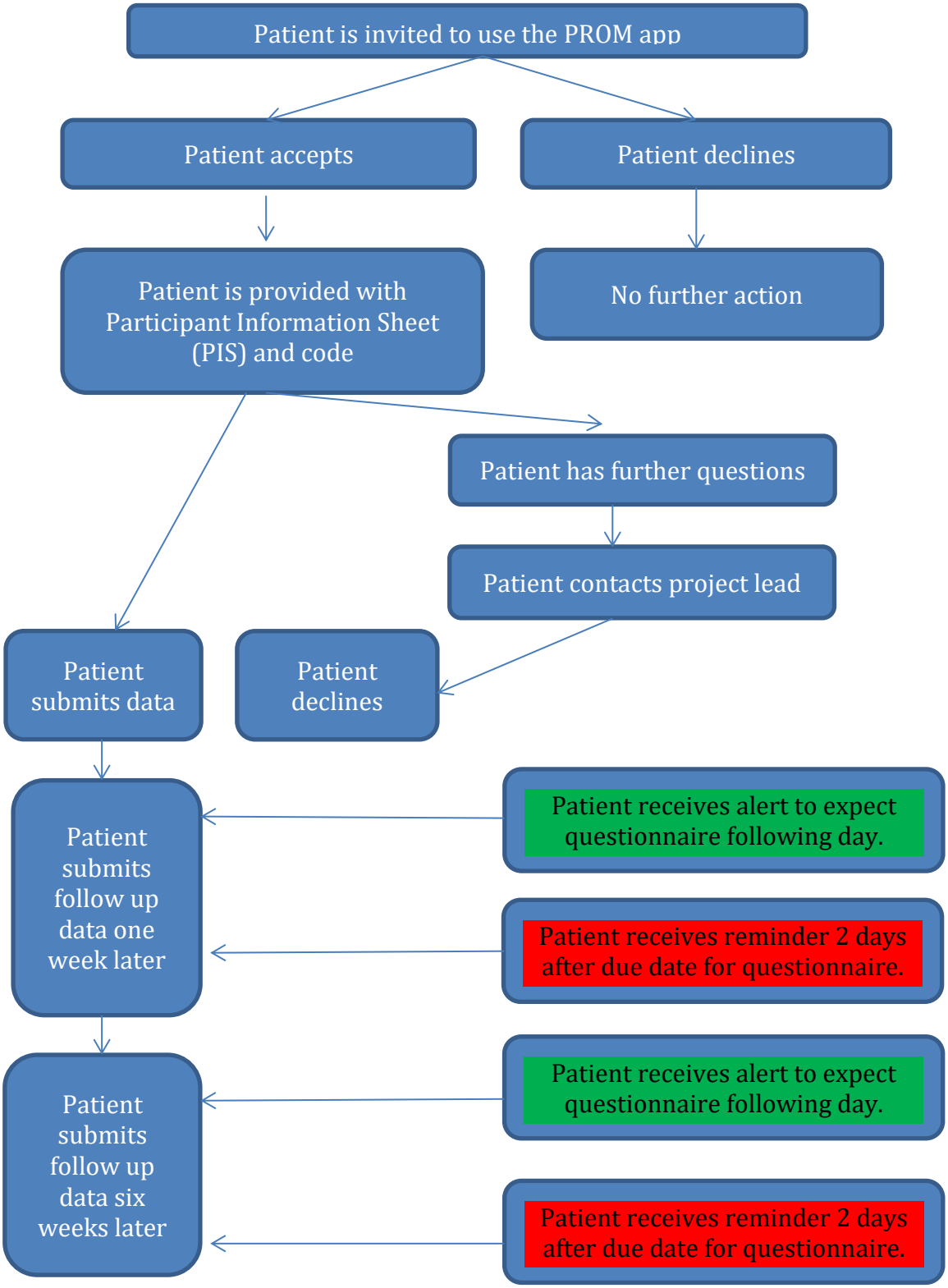


Figure 6.12. The post-pilot data collection process

6.3 Results from the implementation process

Data returns from the implementation process have been limited due to the slow uptake of the app by the profession. Despite concerted efforts during the pilot phase of the app to inform osteopaths about its existence, three years hence some osteopaths are still discovering its existence. This has underlined the long term nature of this project, and the not-inconsiderable task of implementing the app into day-to-day practice. One thing that has become clear is that while outreach visits and social media are at different ends of the communication spectrum, they have been more effective recruitment tools than print or electronic communications in the osteopathic media (*The Osteopath* published by the GOsC, and *Osteopathy Today* published by the Institute of Osteopathy). When considering the long term nature of the project, it is important to look at a more sustained marketing strategy and programme. Marketing is "*the science of meeting the needs of a customer by utilising the expertise of an organisation at the same time as achieving organisational goals*" (Palmer, 2012). A marketing programme is described as "*a coordinated set of activities, thoughtfully created to meeting clear objectives*" (The American Marketing Association, 2016). There are many tools which can be employed to underpin a strong marketing strategy, *e.g.* advertising, branding, websites, conferences, and social media (Twitter and Facebook). This coordinated strategy is based around the "Four Ps":

- Product;
- Price;
- Placement;
- Promotion.

The Four Ps has been revised more recently to reflect the growth of service in marketing and has become the "Seven Ps" to include People, Process, and Physical Evidence (Jobber and Ellis-Chadwick, 2012).

Progress with implementation of the PROM app has been low key. To date (end of October, 2016), a total of 123 osteopaths have enrolled to use the app

(approximately 1% of the profession). The characteristics of the participants are given in Table 6.8.

Table 6.8. Characteristics of osteopaths using the app in clinical practice.

Participants' characteristics	Number of participants
Sex	
Male	n=54
Female	n=69
Region	
Scotland	n=2
Northern Ireland	n=0
Wales	n=1
North West England	n=7
North East England	n=4
Yorkshire	n=15
Midlands	n=13
South West England	n=20
London	n=30
South East England	n=19
Norfolk and Suffolk	n=12
Number of years in practice	
0-10	13
11- 20	45
21-30	39
31-40	17
41 or more	3

The trends in post-pilot use of the app are shown in Figure 6.13. Use of the app began slowly in September, 2015 after a soft launch. Peaks in enquiries tended to coincide with articles in the osteopathic press, the iO annual convention in November, 2015, and meetings for members of the GOsC regional communications network in March, 2016.

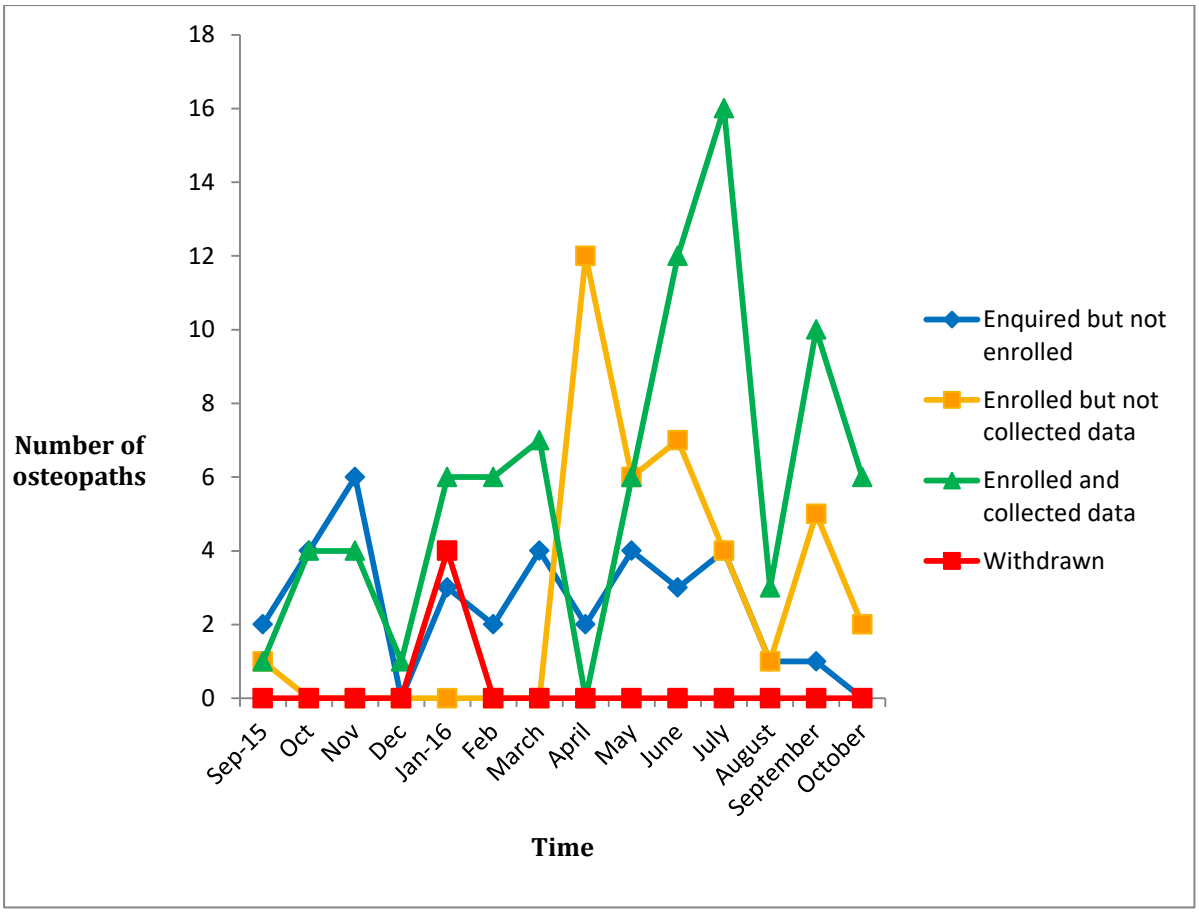


Figure 6.13. Post-pilot data collection participation

In some instances responses to enquiries did not translate into enrolment despite additional gentle follow-up emails. On some occasions osteopaths did respond with comments, for example:

"There's nothing in here for me. This is not what osteopathy is about."

For other osteopaths, it was simply a question of practical issues including moving practice, arrival of babies, taking some time out, or deciding not to proceed with attempting to attain an NHS contract where use of the app had been an implicit part of the planning process. On other occasions (n=8) osteopaths did not respond, while some osteopaths welcomed the reminder and did enrol in the process to use the app. On one occasion, a practice had enrolled to use the app with communications taking place with more junior staff members, only for the practice to be withdrawn by the practice principal. The reason given for withdrawal was

"*I was utterly dismayed to see the narrow focus of this project. This does not capture the full range of osteopathic care, and its holistic nature. I will have to reconsider if I am willing to invest my time in this, and I will discuss this with colleagues in my practice.*"

In other instances, osteopaths who were not using the app were contacted and a range of reasons emerged. These included putting the information pack on a shelf and forgetting about it, reduction in the number of new patients whom to approach, simply forgetting when involved in the consultation process (mainly in single-handed practices), or awaiting logistical changes *e.g.* new staff starting work, or moving premises. Notwithstanding some of these issues, the practices who are using the app are delivering data, and this is shown in Figure 6.14. There is an encouraging amount of baseline data being submitted, but reminders have been sent to osteopaths about asking patients if they have received their follow-up email to encourage them to complete it.

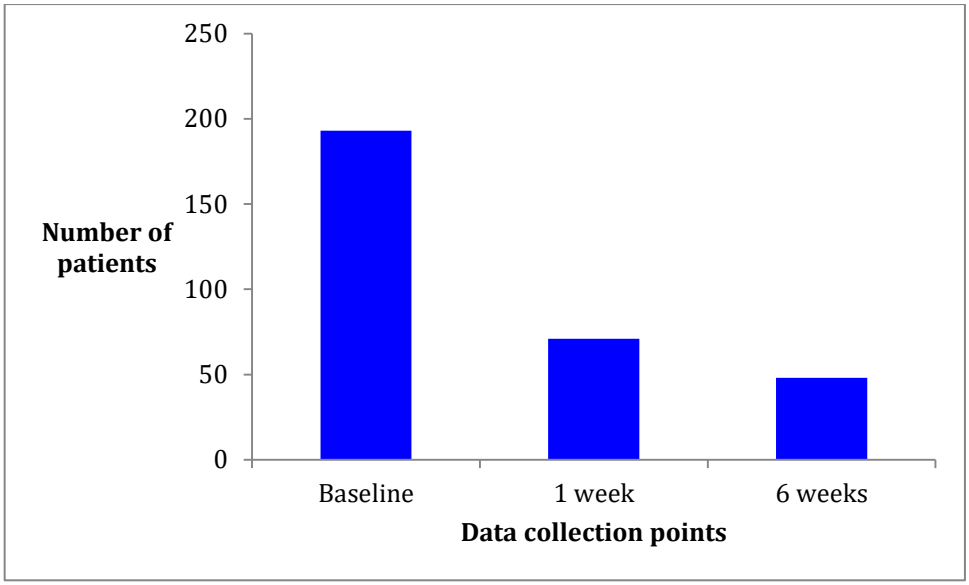


Figure 6.14. Total data collected in post-pilot phase (Sept 2015-October, 2016)

One issue that has arisen is in practices which are using a tablet computer to collect data. Although clear guidance was provided at the start of the data collection

process, in some practices the data have not been transferred from the tablet to allow patients receive their one week follow-up questionnaire and complete it using their own email address. Close attention to the data, and communication with practices where tablets are being used has remedied this issue. This has emphasised the need to closely monitor data returns and maintain close contact with practices while they are introducing the PROM app.

6.4 Discussion

The use of PROMs in clinical practice is becoming increasingly part of the healthcare landscape and the app content development and implementation remains part of a longer term strategy to support osteopaths who wish to engage with this process. However, this process has attracted both interest and criticism: interest for the fact that osteopaths can collect data for their own use, and as a pan-professional initiative, and criticism for the fact that it does not contain PROMs for all joints, all areas of the body, all non-musculoskeletal conditions that some osteopaths treat, and be suitable for paediatric practice. Although this can become slightly frustrating and tedious, it is useful at such times to remember the wise words of Aristotle (684-622BC)

"To avoid criticism say nothing, do nothing, be nothing."

As mentioned earlier, the app content development should be regarded as an iterative process. It will include more PROMs over time, and a small amount of additional information may be added although, mindful of the views of patients, it should not be too long, and should collect information which is relevant to patients. An early barrier which arose was concerned with the fact that "I know my patients are better because they tell me" intimating that a questionnaire, electronic or otherwise, would not add anything of value. However, the role of self-reflection on success of practice and competence in practice has been widely discussed in the literature. Davis *et al.* in their review identified that clinicians had limited ability to accurately self-assess performance (Davis *et al*., 2006). This has been followed by other work which aims to support clinicians in their evaluation of performance (Armson *et al*., 2007; Hawthorn *et al*., 2009; Overton, 2009; Kursurkar *et al.*, 2012).

New regulatory initiatives in many healthcare professions advocate the role of including other clinicians in performance assessment (Andersen *et al*., 2008; MacCarthy *et al*., 2012), but a reluctance exists at national and international levels to involve another obvious person in the evaluation process, namely the patient (European Commission, 2012).

In the early stages of the PROM development the issue of feedback during the treatment process was considered. This was discussed at great length with my supervisors and the funders of the app. The overriding view was that the app would be used to collect robust outcome data, and the greater independence in the data collection process between the clinician and patient the better. One of the main issues associated with this was the fact that not only was PROM data being collected, but also satisfaction and experience data. It was decided that patients would be more frank in their views if the data were being collected independently than being returned to the practice. Although the system lacks the capacity for monitoring patients' progress while treatment is ongoing, patients interviewed as part of the qualitative study (Chapter 2) indicated that just considering some of the questions in the PROM had made them reflect more upon their symptoms, and identify some questions or issues they needed to raise about different aspects of their care *e.g.* self-management initiatives. This increase in patient engagement is a valuable addition to the care process.

Another key challenge in applying the implementation literature to the process is the disagreement concerning what constitutes simple or single and complex/multifaceted interventions. This point has been much debated in recent literature. Common practice favours targeted multifaceted approaches (Bero *et al*., 1998), although the wisdom of this approach has been questioned. Squires *et al.* undertook an overview of twenty-five systematic reviews of implementation approaches to change clinician behaviour (Squires *et al*., 2014). They recognised the barriers commonly cited in the literature, and reviewed the literature to evaluate how effective different strategies are in overcoming such barriers. The evidence for effectiveness of multifaceted interventions was limited and conflicting: some of the conflict has arisen due to methodological shortcomings, predominantly in the

statistical analysis, in the reviews on which the recommendations for multifaceted interventions were made (Wensing and Grol, 1994; Davis *et al*., 1995). When more robust statistical approaches were used in reviews the recommendations for multifaceted interventions are challenged, and suggest that single interventions can be equally as effective (Grimshaw *et al*., 2004; French *et al*., 2010). On completion of their overview of systematic reviews, Squires *et al.* reached the same conclusion having found no compelling evidence to favour multifaceted interventions over single-component interventions (Squires *et al*., 2014).

However, this raises another issue concerning what actually constitutes a single intervention, and a multifaceted intervention. A single intervention can be the introduction of a practice guideline or a computer reminder (Bucknall and Fossum, 2015). A single intervention may be seen as complex by those required to facilitate it, and those clinicians who are the recipients of that intervention. This underlines the flaw in the attempt to create a "one size fits all" recommendation to implementation science. More recent studies on knowledge translation to practice have identified the need for a continuum approach. Harvey and Kitson have developed a new framework to support implementation of new knowledge known as Promoting Action on Research Implementation in Health Services (PARIHS) (Harvey and Kitson, 2015). This approach recognises that a whole range of strategies for facilitation can be helpful; this allows for more intensive facilitation in local settings which are challenging, and a lighter touch in other settings which are more receptive to new innovations or in a state of "readiness to change" (Weiner, 2009). This supports earlier studies highlighting the importance of local leadership and opinion leaders (Harvey *et al.,* 2014), but also the notion of emancipatory approaches in settings who want the freedom to tailor an intervention to enhance its potential for effectiveness (Harvey *et al*., 2002; Seers *et al*., 2012). Using this framework to seek explanatory approaches to implementation instead of seeking causal relationships may be a more fruitful approach to knowledge translation, and introducing new innovations based on that knowledge (Harvey and Kitson, 2015).

In choosing different approaches to the implementation of the PROM app, I have tried to be mindful of the challenges cited in the literature. Even within osteopathic

practice the context can be quite distinct, for example for single-handed, and group practices. This underlines the fact that a continuum of facilitation has been required, and will continue to be required during the longer term implementation of the app. The guidance provided by Grol *et al.* in basic analysis of context has proved invaluable in attempting to introduce what, to many osteopaths, has been and will be a significant change in their practice (Grol *et al.,* 2013). The app pilot process and the initial findings from the implementation stage have been illuminating. It has been clear from the process that although the uptake of PROMs use has been slow it is building especially where third party endorsement has been offered. This has been a notable area of success in the implementation with various osteopaths stating they felt more compelled to engage with the process having heard positive feedback from their peers. It will be beneficial to increase the creation of "champions" to promote the process at arms' length from me. Although outreach visits have been useful to disseminating the findings of the project and increase engagement, their conversion rate to participants in PROM data collection remains questionable. Regular small communications through social media have been useful, and the next obvious stage for engagement to a wider audience through electronic media is a webinar. This process of sharing information has great potential in the future.

When addressing wider engagement with PROMs, an area which is currently lacking nationally is access to information for clinicians who are naïve to their use. I feel there is a level of assumed knowledge associated with the terminology, how PROMs should be administered, and what the findings mean in relation to patient management. This area of education is one which I would include as part of the implementation process through an e-learning package. This would allow clinicians to earn CPD time learning about PROMs, and how to interpret their findings perhaps taking away some of the apprehension that their use engenders anecdotally and which is supported also in the literature.

When introducing any new innovation it is helpful to reflect on the strengths and limitations of an approach, and to identify more systematic approaches to evaluating the implementation of a new initiative. These issues will be discussed in the next section.

6.4.1 Strengths and limitations

Limitations

There are many ways to identify the strengths and limitations of a project: some are systematic while others are based on personal reflection. Some of the issues related to limitations on the implementation of the app include:

i. The PROM contained within the app currently focuses on spinal symptoms, and although other measures of outcome are present *e.g.* patient satisfaction, patient experience, and the transition question. This has clearly been a challenge for some osteopaths and might contribute to understanding why initial enquiries about the app failed to be translated into actual enrolment and participation;

ii. The use of paper-based questionnaires has been requested in some practices where osteopaths have stated that their patients do not use computers. When this question has arisen, I have always asked the osteopaths concerned whether they have asked their patients if they can use a computer, and whether they would be willing to complete a questionnaire online or using their smartphone. On some occasions osteopaths have asked a sample of their patients about computer use and electronic questionnaires, and they have subsequently stated they would give the system a try. In some instances, my question was met with no response at all, and on other occasions osteopaths have been quite adamant that they could speak with certainty on behalf of their patients. Although it would have been nice to offer the dual option of paper and electronic questionnaires, there were a number of reasons why this particular option was not pursued. These included

- The resource implications from using paper questionnaire including the time involved in printing, and distributing them to practices;
- The financial implications of printing, distribution and return postage;
- The financial and time implications of data entry;
- Most significantly, the concern about data entry error.

Although there are a range of initiatives that can be undertaken to try and mitigate against this and identify if it has occurred including checking a random selection of questionnaires for data accuracy, and double data entry, this is beyond the constraints of this study;

- One of the main aims of this study was to develop an electronic system as the availability of use, and capability with electronic devices continues to grow.

iii. The lack of availability of the app to use on Apple devices in the early stage of implementation has been highlighted by both patients and osteopaths. This process has now been completed by Clinvivo.com.;

iv. A final limitation of the implementation phase is that it has been of relatively short duration. As mentioned earlier, there are different processes that can be undertaken to identify whether an innovation is successful, or not. Due to the relatively short period of time for the implementation stage and the challenges of getting osteopaths involved, there is a lack of quantitative data available to assess the project's progress. This will be discussed in greater detail in the section looking at future work.

Strengths

While it is vital to reflect on the limitations of a project on an ongoing basis to identify how it can evolve, improve, and continue to be reflexive to both participating osteopaths and those patients who take the time to be involved, it is useful to examine where strengths lie also. Some of those strengths include:

- Achieving the implementation of the PROM app as a resource for all osteopaths to use. It is important that patients have the opportunity to give feedback, and independently of practices. This helps to demonstrate to patients that their views are important, and their treatment progress matters. This provides a means for the collection of independent outcome data from day-to-day practice for the first time;
- The soft introduction of the app has limited the opportunity to evaluate its progress, but it has allowed more support to be available to osteopaths

while they are beginning to use the app. Recruiting osteopaths to use the app takes time, so it is important that their involvement is retained, and becomes a natural part of their practise. Future initiatives in the app implementation process are discussed in Section 6.4.2;

- Retention of interest for patients and clinicians is vital through what will be a long term intervention. To facilitate this, newsletters are prepared periodically to send to patient and clinician participants from the focus groups and who are involved in the data collection process (Appendix 6.2 and Appendix 6.3 respectively. This information is available also on the NCOR website for any other interested parties and to foster a spirit of openness about the project.

6.4.2 Reflecting on the implementation process

As the implementation process reaches the end of its initial stage, it has been helpful to reflect on its progress. Some aspects of the process have gone well. The information provided to osteopaths to encourage them to join has been suitable with few requiring clarification about what was involved. In a similar manner, the resource pack provided to osteopaths has provided clear information about what to do and useful resources to implement the process into day-to-day practice. From later discussions with non-participating osteopaths, it has been clear that the dissemination process has been less effective with many osteopaths being unaware that the app was available, or some informing me that such a facility would be a good idea. There are limited options for communication with osteopaths *e.g.* two hard copy publications, social media (NCOR's Facebook and Twitter accounts), and talks to regional groups. These options have been utilised, but perhaps the frequency of communication with minor modification of the message on each occasion would have been more helpful. One of the most effective communication strategies was the use of "champions" who spoke independently about using the app, and the fact that it fitted in seamlessly into their day-to-day practice. Marketing and implementation literature provide guidance on the process, and greater use of marketing literature is something I would change in the future. The marketing literature suggests using the type of strategy shown in Figure 6.15 to plan the implementation of a product.

Figure 6.15 Planning for implementation of app.

When considering further the items in Figure 6.16, more detailed consideration can be provided to key aspects of what it is intended developing further the stakeholder analysis undertaken in Section 6.2.

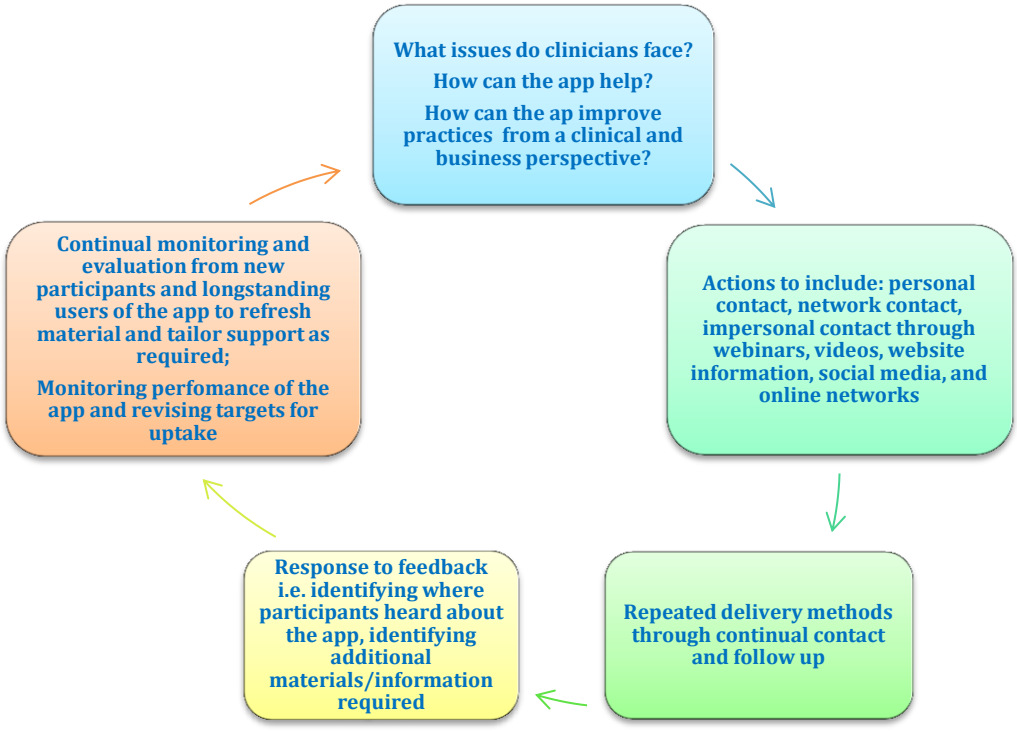


Figure 6.16 Considerations for planning implementation strategy.

A scanning of the current environment for osteopathy involving both internal and external factors by using a PESTLE analysis (Sammut-Bonnici and Galea, 2015).

Political	Political challenges come in many forms. There are those driven from the changing organisation of healthcare and the evidence they require to maintain existing services or commission new ones.
Economic	The pressure to evidence the cost-effectiveness of treatment by external stakeholders e.g. insurers, commissioners, and self-funding patients.
Social	The use of various different forms of healthcare external to the NHS, and the need for greater information about this.
Technological	The increased use of technology in healthcare to acquire and use data within healthcare.
Legal	Compliance with the standards set out by the CAP codes delivered by the ASA; Conforming with the changing requirements of the GOsC.
Environmental	The importance of being able to describe practice to a more informed patient base.

This is supported further by the use of a SWOT analysis (Fine, 2010).

Strengths	Gathering outcome data demonstrating high levels of satisfaction, experience, global and specific change; A simple, user-friendly system with minimal administrative burden for clinicians and time burden for patients; The opportunity to benchmark individual practice against a national profile.
Weaknesses	Limited measurement of outcome initially focussing on musculoskeletal osteopathic practice, and limiting input and engagement from clinicians managing patients with non-musculoskeletal symptoms and/or from paediatric populations; The capacity for developing the app to include other outcome measures is untested at this time.
Opportunities	Gathering outcome data demonstrating high levels of satisfaction, experience, global and specific change. Allowing osteopaths to seek different clinical opportunities *e.g.* commissioned services from primary and secondary care providers, or occupational health departments of businesses; Allowing osteopaths to reflect upon weaker areas of practice and address these through targeted CPD activities.
Threats	Gathering outcome data which demonstrates change which is poorer than anticipated without showing the additional aspects of osteopathic care *e.g.* education and advice. Focussing on musculoskeletal data potentially reinforces the perception that osteopaths deal solely with low back pain; Osteopaths identifying weaknesses in their practice of which they had previously been unaware, and potentially undermining their confidence.

The information from the PESTLE and SWOT analyses provide information to make a more targeted strategy for implementation. Having more time to reflect upon the early stage of the implementation and look dispassionately at what need to be done in the future will try to support the continued use of the app by osteopaths. This process of reflection and will need to be ongoing to ensure that the process is reflexive to changing circumstances affecting the profession in terms of policy and practice in the years ahead.

6.4.3 Future directions

A systematic evaluation of the implementation process is the next natural direction for the study. The literature identifies that new initiatives need to be evaluated to identify:

- whether they have been successful;
- whether they have failed due to inadequacies of the actual intervention concept or theory, a so-called Type III error in implementation science (Dobson and Cook, 1980; Carroll *et al*., 2007);
- whether they have failed due to the inadequate implementation (Rychetnik *et a*l., 2002; Carroll *et al*., 2007).

The processes which can be used to evaluate the implementation of new initiatives are described in the next section.

6.4.4 Evaluation of the implementation process

The challenges of identifying a suitable implementation strategy for a particular setting have been discussed in section 6.1.1, and while the evidence for different strategies is contradictory greater consensus exists on processes to evaluate the implementation of new initiatives. The literature focusses on two main methods for evaluation of implementation:

- Process evaluation (Oakley *et al*., 2006; Rycroft-Malone and Bucknall, 2010;
- Fidelity or integrity studies (Dane and Schneider, 1998; Dusenbury *et al*., 2003; Carroll *et al*., 2007).

Process evaluation

Process evaluation is a method for exploring the implementation, setting, and receipt of a particular intervention (Wight et *al*., 2002). When interventions are distributed over a large number of settings variations are possible in how those interventions are delivered, how they are received, the environment, and all of these factors have potential to change the outcome of that intervention. Arguably a process evaluation should be embedded in all studies to clearly identify what interventions work in particular settings, and for whom (Rycroft-Malone *et al*., 2010).

In this study, the intervention is the introduction of a PROM app into osteopathic practice, the settings are osteopathic practices, and the outcomes are the actual introduction of the app to patients, and patients using the app to submit data. Data for a process evaluation can be both qualitative and/or quantitative (Oakley, 2005). Oakley *et al*., 2006 shared their experiences of the RIPPLE (randomised intervention of peer-led sex education) trial to highlight some of the methodological issues for process evaluation especially when trying to include them within a randomised controlled trial (Oakley, 2006; Stephenson *et al*., 2003). These issues include:

- Process data should be collected from all sites participating in a study;
- Qualitative and quantitative data should be collected;
- Process data should be analysed prior to outcome data to avoid bias in interpretation;
- Statistical processes should be used to avoid or minimise the risk of bias and error in interpreting the study's findings (Oakley *et al*., 2006).

Different methodological approaches and their application to the PROM app process are described in Table 6.9.

Table 6.9. Methodological approaches for use in process evaluation

Qualitative approaches	**Quantitative approaches**
Focus groups with participating osteopaths.	Questionnaire-based survey of all participating osteopaths.
Focus groups with practice-support staff.	Questionnaire survey of practice support staff *e.g.* receptionists involved in implementing the PROM app.
Focus groups with students in in participating Osteopathic Educational Institutions (OEIs).	Questionnaire survey of students in OEIs.
Researcher observation of osteopaths/ receptionists/ students in a practice setting.	Questionnaire survey of patients who have used the PROM app.
Researcher observation of patient completion of the app in practice.	Questionnaire survey of patients who declined to use the PROM app.
Individual interviews with osteopaths/receptionists/students in practices.	
Individual interviews with patients using the PROMs app.	
Individual interviews with patients who declined to use the PROMs app.	

Although the introduction of process evaluations represents an additional layer of cost, activity, and bureaucracy in a study, it offers the potential for greater explanation of an intervention, and its capacity for generalisability amongst a wider and perhaps more diverse populations and settings. As the development of theory-based approaches to evaluations are increasing, they are included more frequently at the developmental stage of clinical trials and other forms of investigations (Stame *et al*., 2002; Power *et al*., 2004; Mars *et al*., 2013). While process evaluations focus on the manner in which an intervention is implemented, the use of fidelity or integrity

studies examine whether the actual intervention was implemented in the manner intended by the developers of that intervention (Dusenbury *et al.*, 2003).

Fidelity studies

Fidelity or integrity are synonymous terms to describe the degree to which interventions are implemented in accordance with the intentions of the interventions' developers (Dusenbury *et al.*, 2003; Dane and Schneider, 1998). Implementation needs to be measured in any intervention as it is known to affect outcomes (Mihalic, 2002; Mihalic, 2004). The concept of fidelity or integrity is defined in the literature as having five key elements as shown in Figure 6.15.

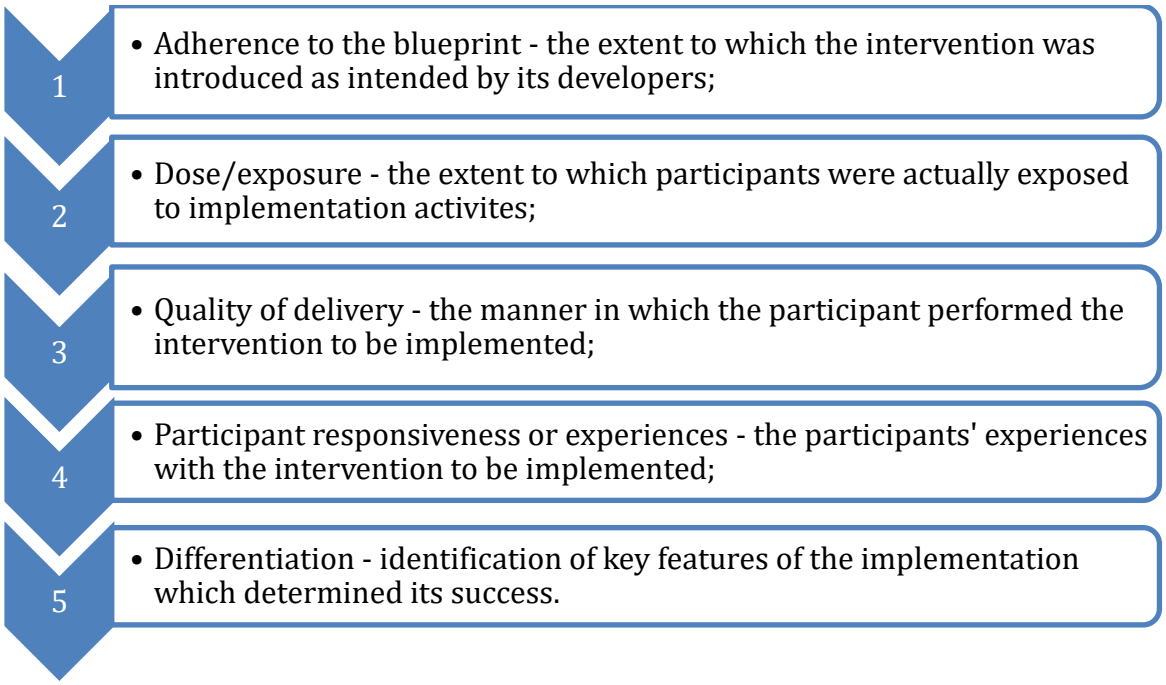


Figure 6.17. The key elements of implementation fidelity (Mihalic, 2002; Mihalic, 2004)

While the literature broadly agrees on these five key elements, there is less agreement concerning how they should be measured. Mihalic argues that fidelity can be measured using any of the five key elements (Mihalic, 2002; Mihalic, 2004), and Dusenbury *et al.*, and Dane and Schneider argue that all five elements need to be measured to present a comprehensive picture (Dusenbury *et al.*, 2003; Dane and Schneider, 1998). However, Carroll *et al.* offer a third conceptual framework asserting that all five elements should be measured and the relationship of each

element to the other should be explained (Carroll *et al*., 2007). This approach could attempt to explain, for example, whether participant antipathy to an intervention has an effect on the completeness and diversity of its implementation. The key elements of the framework are described in Figure 6.17.

Identification of the inter-relationship of each of the components described by Carroll *et al.* provides scope for adapting interventions to local conditions thereby enhancing their generalisability (Grol *et al*., 1998). It has been argued, however, that the need for local adaptation has been overstated, and some interventions simply do not require this (Elliot and Mihalic, 2004). To reach a practical solution, it appears that middle ground must be reached developing interventions which can be easily implemented and with high fidelity, but which also have sufficient flexibility to allow adaptation to particular circumstances where this may be required.

Table 6.10. Key elements in the implementation framework proposed by Carroll *et al., 2007*		
Elements	Elements' sub-categories	Explanation of sub-categories
Adherence		
	▪ Content	Described by Carroll as the "active ingredients" *e.g.* the skills, treatment, knowledge that the intervention seeks to deliver such as using the app in practice.
	▪ Coverage	The extent to which the intervention was delivered, *i.e*. how was the app used.
	▪ Frequency	Relating to the number of times or how often the intervention was delivered.
	▪ Duration	Relating to the length of time the intervention was delivered (short or long term use).

Moderators	▪ Intervention complexity	The level of detail and specific direction of an intervention (clarity of instructions).
	▪ Facilitation strategies	The level to which these are standardised by training and education, with consistent levels of support available to all participants (consistency of materials, and advice/responses to enquiries).
	▪ Quality of delivery	This describes delivering an intervention in a way appropriate to achieving what was intended.
	▪ Participant responsiveness	This relates to the degree to which participants view the relevance of an intervention to them. In the case of the app this relates to how participants regard the use or potential use of the data to their practice.

While different elements of the adherence of implementation are clear, there are many different aspects to what Carroll describes as "moderators". The actual content and description of an intervention may be simple or complex; very clear or too vague to understand. Studies have identified that where interventions are more complex, the greater degree of exactness and detail leads to higher fidelity (Grol *et al.,* 1998). The literature highlights that when interventions are well-planned, with clear components and outcomes identified, they produce higher levels of adherence (Mihalic, 2002; Dusenbury *et al.,* 2003). While complexity can be overcome, it has been shown that simpler interventions are less likely to be stalled by "response barriers" when the manner in which an intervention is received can vary due to the capacity for the intervention to be changed in one of its component parts (Forgatch *et al*., 2005). Carroll *et al.* identified also the importance of support strategies to ensure that training in interventions is uniform, and their consequent implementation will exhibit higher fidelity (Bellg *et al.,* 2004). The role of moderators in the delivery of interventions has been highlighted, and it is important to examine the manner of their working relationship to ensure that dissonance does not exist, and affect the fidelity of an intervention.

Previous studies have highlighted the paucity of work concerned with fidelity of implementation. Carroll *et al*. highlight that too frequently fidelity studies have focussed on the measurement on one aspect of the process, namely adherence, to the exclusion of other equally important aspects (Resnick *et al.,* 2003; Penuel and Means, 2004; McGrew and Griss, 2005). The proposal for the third framework is to ensure that a more inclusive approach is used which attempts to identify the multifactorial nature of an implementation, and the achievement of high fidelity which will ensure successful implementation (Nutley and Homel, 2006; Carroll *et al*., 2007). The fidelity studies raise many areas for consideration in implementing the PROM app, and highlight good practice for successful implementation. The strength of the intervention is that it is relatively straightforward, the potential area for low fidelity is its further implementation by individual practitioners as they embed use of the app into their local settings.

6.5 Conclusions

The implementation process described in this chapter has been challenging. Recourse to the literature provides limited guidance since there is a lack of agreement concerning terminology, and descriptions of interventions are frequently lacking in detail (Goeschel *et al*., 2012; Moore *et al*., 2015). A lack of clear reporting guidelines of implementation studies perpetuates some of the recognised issues (Pinnock *et al*., 2015). Guidelines to support uniformity of reporting will be welcome allowing novice researchers to implement well evidence-based interventions, and report their findings to a recognised standard for publication.

The implementation stage continues to be a learning process for myself and the profession. The participants using the PROM app, both osteopaths and patients, have been generous with their time in implementing the app in practice, and in giving their feedback to help to improve the process as it continues to develop.

The process evaluation will be helpful in identifying further strengths and areas where more attention and development is required. The main lesson is that this is an iterative process and not a completed study. It was the intention for this study at the outset that it should continue to develop by introducing further PROMs and including a wider patient population. The challenge will be to deliver this in a timely manner while retaining the quality of the process, and the quality of the data collected.

7.

Discussion

This thesis describes the development, evaluation and initial implementation of a national programme for the use and collation of patient reported outcome measures (PROMs) in osteopathic back pain services in the UK. In this final chapter I will describe in brief the findings of my research, a critical appraisal of its strengths and limitations including the extent to which that research met my original aims and objectives, and what implications my research holds for future research, policy, and practice.

7.1 Summary of main findings

The first part of this PhD was to explore the utility of creating an electronic data capture system to gather Patient Reported Outcome Measurement (PROMs) data for UK osteopaths. I explored the beliefs and views that clinicians (osteopaths, chiropractors and physiotherapists) and patients had about using PROMs, and found that patients were willing to complete PROMs, but clinicians reported a range of views from wholehearted enthusiasm to deep scepticism. Variation in knowledge and experience was considerable between and within professional groups, and enthusiasm for PROM use tended to increase with familiarity of application. I concluded that it was vital to engage clinicians in the process of development and implementation of any patient reported outcome measurement system to ensure that it is used in practice.

In my initial scoping of PROMs in musculoskeletal care I found there are a large number of PROMs: the settings for which they have been developed, and their

measurement properties vary considerably. To ensure that data collected by healthcare professionals are useful it is necessary to use a PROM which is appropriate for an acknowledged setting, has the capacity to measure change in an identified population, and has good responsiveness and reliability also. One such appropriate PROM is the Bournemouth Questionnaire (BQ), so I did a systematic review of its measurement properties using the COSMIN tool. Although this tool is being used increasingly to evaluate PROMs it was not without its challenges. Overall, the BQ was found to perform well with differing performance ratings for published studies addressing reliability, responsiveness, translation and cross-cultural validity.

Using the findings of the systematic review, and the two qualitative studies, I developed the content of the PROM app which was transformed into functioning web and mobile apps (Clinvivo.com). The development and piloting of the PROM app showed that the app performed well requiring only minor modifications during the pilot process. Ensuring patients comply with submitting longer term follow up data (at 6 weeks) was challenging. The quality of data submitted was high, and this allowed a range of statistical analysis from the included PROMs.

Early findings of the implementation of the PROM app showed that educational material in paper media, and outreach visits have their place in an overarching strategy, but third party endorsement was the most notable factor to influence implementation of the app. As the implementation process continues I describe a process evaluation to be undertaken to evaluate the implementation strategy.

7.2 Critical appraisal, strengths, and limitations

At the beginning of this PhD, I identified aims and objectives for this body of work. They were:

Aims:

I To design and develop a nationwide system of collecting routine PROM data from osteopathic patients;

II To enable the establishment of baseline standards for outcomes for patients presenting with musculoskeletal conditions to osteopaths as a benchmark comparator.

In order to meet the aims of the study, a series of research objectives were identified:

i. Undertaking a review of the literature concerning the use of PROMs in clinical practice, and the different formats in which they have been used;

ii. Conducting qualitative work to identify the views of patients about the concept of data collection in clinical practice, and their views concerning three different patient reported outcome measures (PROMs);

iii. Conducting qualitative work to identify the views of osteopaths on the concepts of data collection in clinical practice;

iv. Conducting qualitative work to identify the views (and experiences) of physiotherapists and chiropractors concerning PROMs and their use in a clinical setting.

v. Scoping, and systematically reviewing the literature concerning the measurement properties of a selection of key PROMs;

vi. Using the review and qualitative findings to develop content for an app suitable for use via the Internet, mobile telephone or other mobile device e.g. a Tablet computer;

vii. Pilot testing the app to assess its functionality, the feasibility of using the app in clinical practice, and the clinimetric performance of the PROMs in an electronic format;

viii. Examination of the responsiveness of the PROMs in UK osteopathic clinical settings and identify baseline standards for patients attending with musculoskeletal symptoms;

ix. Examination of the test-retest reliability of the PROMs in UK osteopathic clinical settings and identify baseline standards for patients attending with musculoskeletal symptoms;

x. Examination of data concerning patient satisfaction and experience in clinical practice;

xi. Refinement of the app based on feedback, and its implementation into day-to-day osteopathic practice.

I achieved my original aim of developing a nationwide PROM data collection system for osteopaths. A small dataset of outcomes (n=404) had been collected at the end of the PhD process, and data collection continues to date. Over time the dataset will continue to grow allowing greater capacity for benchmarking across the profession. The concept of benchmarking has evolved, and it has become an increasingly accepted part of healthcare since the 1990s (Benson, 1994). There are four kinds of commonly accepted types of benchmarking: internal, competitive, functional, and generic. Benchmarking in osteopathic practice can be internal within practice, competitive between practices, functional as it allows osteopaths to reflect on their data for identifying CPD opportunities, and generic as it allows osteopathic care to be explored in relation to other therapeutic approaches. Ideally benchmarking needs to be a team or pan-professional process, as the outcomes identified may involve change to current practices and/or management with effects felt by both clinical and support staff (Mahlknecht *et al.,* 2016). Qualitative data from clinicians who had experience of using PROMs in their practices noted the need to engage all clinicians and support staff in this process. This attempted to ensure using PROMs fulfilled a useful purpose and did not become another bureaucratic burden with little value as some had suspected initially.

7.2.1 Qualitative research

The qualitative study involving patients was an invaluable starting point for the development of the app and its content development. It informed whether the development of an app *per se* was useful, the type and format of its content, and an acceptable completion time. Although consulting patients is an important part of research, it is still too frequently neglected. This study made patients the focus of the research. It filled a useful niche in qualitative literature but it has placed electronic PROM completion in context among other studies in this area where findings are contradictory (Roberts *et al.,* 2014; Jenkins *et al.,* 2016; Malhotra *et al.,* 2016).

The qualitative work involving osteopathic patients could have benefited from the inclusion of more new patients and especially those in a more “acute phase” of their symptom presentation. This would have provided further information concerning

how patients in new and perhaps extreme pain states would have regarded the request for PROM data collection in addition to coping with their symptoms. The opportunity to interview clinicians from three different professional groups with varied experiences of using PROMs in clinical practices provided a useful complement to the data provided by patients.

Although the broad spread of views concerning the use of PROMs was a strength, the lack of balance in numbers between the professional groups was a weakness. The results could have been enhanced if more physiotherapists were included: their numbers were small and greater recruitment would have potentially added to the rich data provided by osteopaths and chiropractors. Conducting interviews or focus groups with osteopaths using a non-osteopath as the interviewer might have produced richer and perhaps more contentious data from among the profession.

7.2.2 App content development and pilot

The development of content for the app was challenging. Input from patients stressed the importance of creating an app which was quick and easy to complete. This, by necessity, limited the amount of data that could be collected and used by individual clinicians and the profession at large meaning the utility of data for national policies and guidance informing local practice initiatives, staffing, need for facilities, and training is reduced.

The creation of the app by Clinvivo.com, and the content into a useable app has been a key strength of this study, it still has considerable scope for future development. For example satisfaction and experience are multifactorial concepts, ideally separate questionnaires would be included within the app to explore these areas in much greater depth, but this has to be placed within the context of completion time and potential patient burden. Incomplete data from the app has limited value so the tension remains between acquiring completed data sets which have some limitation on their content with more data contributed by far fewer patients who may be less representative of patients attending osteopathic practices.

In the test-retest reliability strand I attempted to embed the process within day-to-day practice but this was not particularly successful. Although the stability of the population could be evaluated by examining patients' responses to a question on global change, a preferred method in hindsight would have been to request patients to complete a questionnaire one week prior to their appointment, and then again on the day of their appointment prior to any treatment taking place (Bolton and Breen, 1991). This would have included a greater administrative burden on practices but would have potentially yielded more useful data.

The numbers of participants in the responsiveness strand of the study also represent a significant limitation of the study. Although baseline data were submitted in good quantities, the lack of follow up data at six weeks represented a challenge for evaluation of measurement properties. Since recruitment was through third parties there was no capacity for patients to be sent any additional reminders to those dispatched through the automated facility in the app. As a result of the low numbers in the responsiveness strand of the study, recruitment will continue outside of the PhD to ensure that an increased amount of data is collected compatible with the guidance of the COSMIN group *i.e.* 30-50 patients.

7.2.3 Implementation

The implementation process was in the very early stages at the end of the PhD period but it clearly required a more targeted and strategic approach over the long term to support greater uptake among osteopaths: uptake of the app by osteopaths during the first nine months of its availability amounted to 2.7% of the profession. The role of the app in benchmarking was identified as an aim for this PhD, but to be successful the benchmarking process requires feedback on a regular basis to participating osteopaths to allow the quality of care to be maintained or improved where necessary (Peabody *et al.,* 2016). This may be through adherence to clinical guidelines, process changes, or more patient-focussed management (Meissner *et al.,* 2006; Rossignol *et al.,* 2011). To support practice and personal reflection, osteopaths may be supported in future by a national dataset with published information concerning outcomes and success factors e.g. patient experiences, and patient satisfaction potentially producing higher levels of performance (Bayney, 2005;

Collins-Fulea *et al.,* 2005; Greenough, 2006; Ellershaw *et al.,* 2008; Meissner *et al.,* 2008).

The app focussed on a narrow area of practice, namely the lumbar spine. Although earlier survey work has indicated that the lumbar spine is the most frequent area of symptom presentation, there are many other body areas and symptoms that osteopaths encounter in their day-to-day practice (Fawkes *et al.,* 2012). The lack of content for clinicians who focus on paediatric practice has reduced the implementation to a wider audience, and drawn criticism that the app is not "osteopathic or holistic enough".

Face-to-face encounters and the use of champions have yielded an increase in uptake of the app but with limited resources this happens at a much slower pace than is desirable. The lack of champions is an issue due to the newness of the project and this will increase over time.

7.3 Implications of the research in this thesis

This research addressed some important issues concerned with gathering patient data in clinical practice. Although the introduction of electronic clinic notes is becoming increasingly common, there are still barriers to overcome in the collection of data concerning patients' assessments of their outcomes of care. The literature identifies some of these issues for clinicians, but the patient is all-too-frequently ignored. When proposing any new innovation it is important to engage patients in the development process. Their input not only contributes to the success, or lack, of an innovation but also provides guidance for implementation, and long term sustainability.

7.3.1 Patient-centred care

The role of the patient has changed in healthcare over the past 30 years, but their input is still limited in some important areas of practice, and changes can still be imposed instead of engaging patients in the change process (Darzi, 2008). Although patients are becoming involved increasingly in the development of outcome measures to ensure that changes important to them are evaluated, their views

concerning the manner in which such data are collected can be neglected (Hills *et al*., 2016; Jenkins *et al*., 2016; Malhotra *et al.,* 2016). This stance persists in spite of the increased technological innovation and coverage of technology across the population in the UK (Longley and Singleton, 2008; Ofcom, 2015). The assumption by some clinicians that patients may be unwilling, too busy, lacking IT access or capability potentially limits a wealth of valued information for clinicians in evaluating their performance. This parentalist approach potentially reduces compliance with clinicians' recommendations (McAuley *et al.,* 2014; Valier *et al.,* 2014). Although some patients may prefer such parentalistic management, I would suggest this is becoming less prevalent, and patient management is becoming more negotiated over time (McIver, 1991a; Thomson and Sunol, 1995; Leach *et al.,* 2013). This expectation to be consulted was identified by patients in my qualitative study.

For many of the patients receiving osteopathic care over a sustained period, and whose treatment had begun as a last resort, being able to contribute feedback was important. Some patients expressed surprise that this had not happened earlier in osteopathic practice, while others recognised the value of evidence in discussion with healthcare providers e.g. their general practitioners, or for increasing access to osteopathic care through the NHS. Employing a PROM which would not identify meaningful change within an osteopathic primary care setting could potentially undermine the value of a data collection system and the data generated. Using a qualitative approach to explore patients' views in depth was invaluable at the start of the app content development.

The findings of the qualitative study have wider application to both internal and external policy in the profession and for, research and clinical practice. Patients clearly voiced their willingness to contribute feedback in their interviews, and this has been supported further in work commissioned by the GOsC (GOsC, 2014). Policy for the profession could now encourage the use of patient feedback on a regular basis from which osteopaths can learn from reflection upon their practice. The use of independent data collected by a third party also holds value for marketing practices making it more explicit to patients about osteopaths areas of special interests, and ratings of experience and satisfaction from patients. When considering the wider

healthcare arena, patients noted that their PROMs data could be used to communicate with other healthcare professionals. This drive towards a more standardised "language" using data collected at common time points could assist osteopaths in engaging with other professional groups *e.g.* commissioner, while reassuring others e.g. health insurance providers, that their patients are receiving effective care with which they are satisfied.

When considering the findings of this study with respect to research, it would be valuable to engage patients who participated in the qualitative study to identify whether they had been offered to complete PROMs at the practices they attended and whether their actual experience matched what they perceived. Some patients suggested that it raised questions about their about symptoms, self-management, and items they wished to discuss with their osteopath: further qualitative work would identify if using PROMs had had an effect on any of these considerations.

Implementing PROMs in clinical practice can be challenging especially for small or single-handed practices. This research identified that clinicians should not feel apprehensive about asking patients to complete PROMs. Many will welcome the opportunity, and support is available for patients who may hold greater willingness than confidence in their capability to use IT. There are lessons for some clinicians in becoming more familiar with aspects of IT themselves so they can respond to patients' questions or support them in using an app (web or mobile): for others there will be value in learning more about PROMs and what the findings actually mean for patient and practice management. The next section will explore the views of clinicians in greater detail.

7.3.2 Clinician engagement

The opportunity to interview clinicians from three different professional groups with varied experiences of using PROMs in clinical practices was illuminating. It identified issues for the PROM pilot, and informed a range of practical issues in the implementation of the study e.g. assurances concerning the use of the data, making the process straightforward with minimal practice and patient burden.

The contrasting views between different professional groups were surprising and in some ways provided reassurance about the actual process of using PROMs in clinical practice. It suggested clinicians increasingly valued their use with exposure and experience, and gaining feedback about practice and personal performance was seen as a worthwhile investment of time. The extent of the parentalist views in one of the professional groups was surprising, and highlighting this factor has been useful when discussing PROMs with clinicians who are initially reluctant on behalf of their patients to engage in using PROMs.

These findings have important implications for policy and practice. Increasingly patients will be asked to provide feedback within healthcare settings, and osteopathic practice should not be any different. The role of the regulator in requiring more evidence of gathering patient feedback concerning outcomes of care, quality of practice, communication is not to be underestimated. The Osteopathic Practice Standards (GOsC, 2012) requires osteopaths to demonstrate standards of care which meet four key areas, namely "Communication and patient partnership", Knowledge, skills and performance", Safety and quality in practice", and "Professionalism": within each of these areas the use of PROMs and other outcomes collected using the PROM app has been suggested GOsC, 2017). This is likely to be reinforced by the profession's professional association (the Institute of Osteopathy), and ultimately as part of new CPD proposals by the regulator (GOsC, 2017). As mentioned in the previous section, increasingly external stakeholders such as insurers, or within funded services commissioned by the NHS will require PROM data from osteopaths if they wish to be included in these care streams. In view of the new changes being introduced within the NHS as part of management of back pain, familiarity with using PROMs and the ability to interpret their findings will make osteopaths more attractive providers of care (Greenough, 2017).

In clinical practice promoting the potential value of PROMs from the perspective of broad outcomes such as experience and satisfaction has been helpful as clinicians were often more receptive to these softer measures of change. The fact that such measures could feed into a process of practice-based audit allowing greater exploration of what makes a good or poor experience or satisfying clinical encounter

has great potential value particularly in view of changing regulatory requirements. Although data collection may be becoming more commonplace, assurances on confidentiality and use of data must be provided for patients and clinicians. Data should be used to support patient and practice management, with any further use of data being made explicit to the patient at the outset.

Although the value of PROMs in terms of engagement and communication has been cited in the literature, their impact on clinical care is still unexplored. Focus on performance through PROMs' use can elevate vigilance in practice, this can lead to referrals for symptoms which might not have otherwise been detected, better notation of consultations, and discussion around symptom management. Whether this actually translates into changes in patient management has been unexplored. This is an area for future research in osteopathic practice. Further qualitative interviews with osteopaths who have successfully been using PROMs in practice will help to identify if they feel their practice and patient management has changed. The introduction of an educational package to support osteopaths and other clinicians in the use of PROMs is another important area of research. Innovations may be introduced with lack of clinician engagement and support as if new skills can somehow be acquired through osmosis. A tailored package could be tested to identify if information concerning feedback and interpretation of findings has a beneficial effect of patient care and/or clinician confidence.

7.3.3 Questionnaire measurement properties

The systematic review of the measurement properties of the Bournemouth Questionnaire (BQ) is a new addition to the PROM literature. Although the VAS and the RMDQ have had their measurement properties assessed in other publications, the BQ has not undergone systematic review using the COSMIN tool prior to this study (Miller *et al.,* 2008; Gupta *et al.,* 2014; Chiarotto *et al.,* 2016). Although the science around measurement properties is not new, it is becoming more organised in terms of agreement on terminology, and concerning what concepts are actually being evaluated on statistical analysis (Mokkink *et al.,* 2010; Terwee *et al.,* 2012). The COSMIN group have played an implicit part of this process, and it is for this reason that the COSMIN tool was selected (Mokkink *et al.,* 2006).

The BQ, developed for a private practice and primary care setting, is being used increasingly by clinicians in osteopathy, chiropractic, and physiotherapy. As its use becomes more widespread and translations increase in number, the opportunity to evaluate its measurement properties becomes greater. Although I do not wish to repeat my reservations about some of the content of the COSMIN tool, I felt that the conclusions from using the tool undermined the value of the BQ. In a similar manner to using critical appraisal tools to evaluate research studies, contemporary standards for reporting are being applied to studies which may have been published two or more decades ago. The same can be said for using the COSMIN tool when evaluating the measurement properties of PROMs originally developed 10-20 years ago (Bolton and Breen, 1999).

Although the COSMIN tool highlights areas of value in both the neck and low back versions, there are other areas where evaluation is too subjective in my view. The convention for choosing the lowest "denominator" of assessment can produce a more negative view of a study which can be evaluated as excellent in many respects but does not meet current criteria. Notwithstanding the cautions about using the COSMIN tool, evaluating all of the BQ studies has been a useful endeavour and has reaffirmed confidence in its inclusion in the PROM app to collect data in an osteopathic setting. In addition, while the ODI and RMDQ have been recommended as useful PROMs for surgical practice and in secondary care, their use in osteopathic practice was not favoured by patients in qualitative feedback and less appropriate for patients commonly attending osteopathic practices (Bombardier *et al.,* 2000; Harms *et al.,* 2010). The BQ facilitates data collection for use in benchmarking practice, and will encourage interested osteopaths to reflect on their practice in relation to the data from their peers.

The COSMIN checklist is challenging to use, and may be particularly daunting for clinicians with limited experience in evaluating measurement properties. Although the EMPRO tool has failed to be developed and implemented in research practice, a new tool has been proposed which is aimed specifically at clinicians and early researchers who may find the complexity of the COSMIN tool very daunting (Valderas *et al.,* 2008; Francis *et al.,* 2016; Maratia *et al.,* 2016). This development

has been greeted with interest by the COSMIN group although they have expressed views on the shortcomings of the new tool (Terwee *et al.,* 2016). The COSMIN tool is likely to be used in research investigating the measurement properties of different instruments in the future, however its complexity and subjectiveness in some areas may raise questions concerning the conclusions drawn from its use.

7.3.4 Collecting patient reported outcomes

The patient reported outcome questionnaire we developed for the app contains 16 questions at baseline, but there are some additional questions that it might have been useful to add, which would benefit clinical practice, and potentially signpost future research. These questions focus on the ongoing collection of adverse events data. Discussions have been held with a range of stakeholders, and other researchers concerning the merits of this approach where there are a number of challenges. Significantly classification remains problematic when applying information from a research setting to a clinical setting. Although work has been undertaken in this area by Carnes et al., involving a range of stakeholders, a further development of this work would involve greater numbers of patients to identify their views on what is considered "adverse" and what is anticipated post-treatment soreness (Carnes *et al*., 2010). Other views were expressed anecdotally that presenting patients with a list of potential symptoms invites an unjustified focus upon them. The use of an electronic data capture system would appear to be well-disposed to collecting adverse events data if a sound and reliable approach to data collection could be found. Alternatively, a free text area could be provided initially in the app's follow up questions to allow patients to submit their reactions. This, in turn, may allow a classification system to be developed which would allow ongoing data concerning treatment effects perceived as both positive and negative by patients.

The descriptive data collected during the PROM app pilot have provided updated information to that collected in the standardised data collection study in 2009 (Fawkes *et al.,* 2012). This process of updating the findings from 2009 allows the professions to view how it has changed in terms of the patient profiles attending for treatment, and indicating whether patients are attending with different symptom

pictures. This may indicate niche areas or populations for future research which have previously been overlooked.

Further, data collected during the pilot of the PROM app focussed on evaluating the measurement properties of three PROMs (VAS, BQ, and RMDQ) when used in an electronic format. This approach has immense value due to the increasing use of electronic media in healthcare management, and increasing access and capability in using information technology. Although the number of completed sets of outcome data is small for all PROMs, evaluation of the measurement properties was undertaken. This evaluation was not without its challenges due to the approaches advocated by different research groups, and the apparent interpretation of their approach. It is important to distinguish between evaluation of measurement properties, and the interpretation of those findings. De Vet *et al.* stress the importance of the interpretation of measurement properties for a PROM within the context of a particular population and setting (de Vet *et al.,* 2010). The use of MIC values is considered at the individual level although they are determined using data from groups of patients. Inherent in this interpretation are different forms of uncertainty, and it is for this reason that information concerning the distribution of change scores becomes extremely important. Although further work concerning evaluation of the measurement properties of the BQ, RMDQ, and VAS in electronic formats is required using larger data sets and with careful methodological consideration, this work represents a starting point.

The pilot study confirmed that the app is feasible to use and returns well-completed data. Future challenges for clinical practice will be to encourage greater uptake among the profession to obtain a more representative set of data reflecting the different "styles" of osteopathy. There is much to be learned from those osteopaths who have become "super recruiters" in managing to engage patients to complete PROMs at all stages of data collection: this is an area of potential interest for all researchers due to the challenges of patient recruitment. The creation of a virtual hub to allow osteopaths to share best practice would be a valuable addition to CPD opportunities for osteopaths wishing to collect patient feedback. The benefits of the data for use in profiling practices while observing the Committee of Advertising

Practice (CAP) code outlined by the Advertising Standards Authority (ASA) must be emphasised more strongly. At the present time it is hard for osteopaths to see the benefits of participation at the early stage of this process. This should change over time as the value of the data becomes clearer, and that data is used in a more widespread manner.

When considering the effect of the PROM app on policy, there are several areas where potentially it has great merit. The data collected are useful to third parties wishing to know more about the profession generally, and about individual osteopaths also. The type of data collected including duration of symptoms, and the change scores achieved has the potential to demonstrate the saving to publicly-funded services. Patients in osteopathic practice(as in private chiropractic and physiotherapy practice) can self-refer access care more rapidly, with less associated bureaucracy, placing less strain on GP practices, and reducing the slide into chronic pain states with the costs associated with those both to the individual public funding. The collection of outcome data including useful PROM data, with high ratings for patient experience, and patient satisfaction demonstrate also to insurers the potential that osteopathic care offers and the merit for its inclusion as a provider benefit for insurance subscribers.

7.3.5 Implementing new technologies and ideas

When investigating research concerning the implementation of new innovations or interventions, it is striking how different the approaches can be and how poorly they are described and evaluated. When attempting to implement the use of the PROM app into clinical practice, guidance came in many forms *e.g.* guidelines, reviews of implementation studies, and findings of process evaluations. In addition, basic marketing information provided some guidance on how to make the PROM app appealing, how it could be branded to support osteopathic practice, and the importance of "champions" who could act as independent voices to endorse the app. From a research perspective, the implementation phase indicates the need for further good quality trials comparing different implementation processes. Clarity of terminology is at the heart of the research process to ensure that complex implementation processes are well-described, while single interventions are

accurately described. In too many studies at present, complex interventions are described as single interventions limiting the application of such research to practice.

The implementation phase was particularly challenging as much of the advice on implementation activities focussed on resources of time, money and staff: all of which were naturally limited at this stage of the project. There have been considerable lessons learned as the implementation phase has progressed. At a practice level, needs expressed for ongoing clinician support have been surprising in a streamlined system developed for busy practices and clinicians. Key areas for support have been guidance on how to introduce the topic of PROMs within an initial consultation where patients may be attending for the first time, or returning with a new episode after an absence. Finding a helpful form of words which act as an encouragement to complete the PROM app while not being coercive has been challenging. This has meant that more one-to-one support has been required, and additional resources have been created to facilitate the data collection process in practice. Gentle reminders to clinicians have also been frequent since PROM information packs often have arrived at practices and, despite initial enthusiasm for the concept, have been consigned to desk drawers or nearby shelves. Regular follow up contact has been introduced into the process to encourage the start of data collection in a timely manner.

One of the more disappointing aspects to the implementation phase has been the lack of engagement by clinic staff at some of the osteopathic educational institutions (OEIs). Pre-registration students have very crowded schedules but they are being prepared for the realities of practice life, and increasingly this reality requires evaluation of patient feedback and symptom change. The lack of introduction for students to PROMs and other forms of patient reported evaluation is a concern for students and the profession in years to come. Continued engagement with clinic staff to encourage the use of PROM data collection will be required over the longer term. This is an important aspect of educational policy for osteopathic education. Manual therapy clinicians in different disciplines are emerging from preregistration training with familiarity with PROMs, and osteopaths face the possibility of being less well prepared for practice life.

7.4 Future research

7.4.1 Patient reported outcomes for children

One of the most challenging areas for osteopathic practice is the evidence base for paediatric care. While anecdotally this area attracts high reported satisfaction and benefits, this has failed to be translated into research findings. Increasingly there are demands from external agencies for underpinning evidence to support claims (ASA, 2016a; ASA, 2016b). Currently a range of different initiatives are underway to add some clarity to the challenges of describing practice in the absence of clinical trial evidence. At the heart of this challenge is the need for clarity about what constitutes evidence as discussed in Chapter 1 (Rawlins, 2008). One initial endeavour in osteopathy is to encourage the use of ongoing data collection in clinical practice.

Although it is difficult to establish the number of osteopaths who treat children, data from the 2009 Standardised Data Collection exercise discovered that 9% of the consultations recorded involved children (Fawkes *et al.,* 2012b). Describing paediatric practice is challenging due to the paucity of available data. In a 2015 data collection study supported by the Sutherland Cranial College of Osteopathy, the training organisation for cranial osteopaths, Wilkinson *et al.* reported that 13% of presentations involved unsettled babies/infantile colic. A total of 14.7% of the patient sample (n= 530) were under two years of age, and 13% were aged under one year (Wilkinson *et al.,* 2015).

The challenge in collecting paediatric practice data is to have a system with a generic measure with utility for the varied range of clinical presentations, and having the capacity to yield a suitable condition-specific measure based on the symptoms reported on initial data submission. Even within specific conditions there can be a variety of different measures which evaluate a specific aspect of a condition e.g. in patients with cerebral palsy changes in gait, pain, and sleep might need to be assessed to identify if treatment is producing an effect.

7.4.2 Collecting data on non–spinal musculoskeletal complaints

Analysis of data from the pilot identified that patients reported additional symptoms in a number of other joints in addition to those in the spine. When excluding the

patients experiencing referred symptoms, the most frequently reported areas of symptoms in peripheral joints are shown in Figure 7.1.

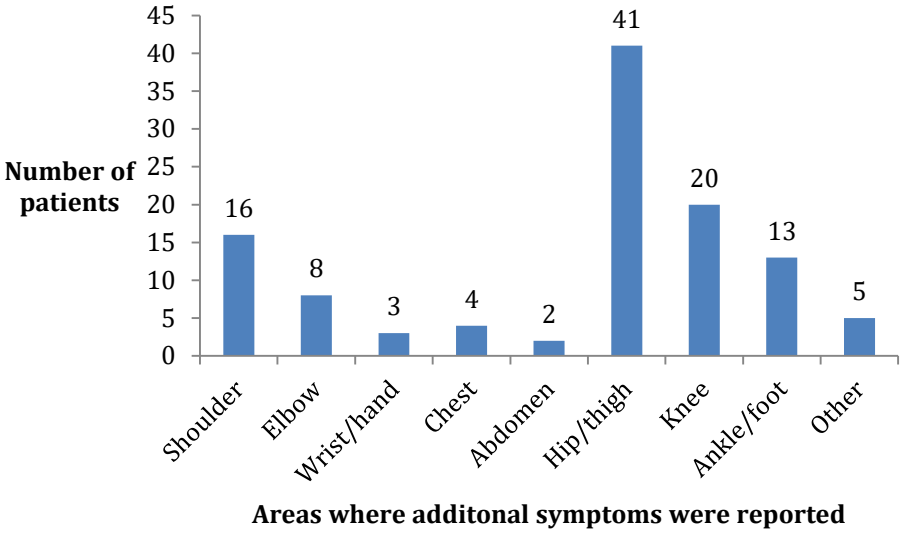


Figure 7.1 Symptomatic peripheral joints identified by patients

While the three most commonly reported symptomatic peripheral joints have been identified, identifying specific PROMs suitable for primary care practice may be more challenging due to the range of complaints, questionnaires and need for quality appraisal to identify the appropriateness of these questionnaires. There are challenges between identifying measures which have utility for external funders while being valid, reliable and responsive within an osteopathic setting. This will involve considerable further investigation and evaluation.

7.5 Conclusions

The use of continuous outcome measurement in osteopathic practice can help to identify changes in patients' symptoms. Monitoring patient outcomes is a reality for osteopaths as they provide healthcare in the 21st century.

The qualitative studies involving patients and clinicians has identified that PROMs should be relevant and fairly quick and easy to complete; the use of electronic data capture is feasible and appropriate with patients even in instances when they may need support. Clear information concerning the limits to the use of their data, particularly their restriction on data being shared with large commercial enterprises

e.g. pharmaceutical companies is required. The use of PROMs in the consultation process should not disrupt the flow of the consultation, and distract from the focus of the consultation i.e. care of the patient.

For clinicians PROMs can be used as a form of standardised language to improve communication with other healthcare professionals and external stakeholders, views were expressed that this may improve access to the choice of services in the management of low back pain by publicly-funded healthcare.

Meeting patients', individual clinicians', and other stakeholders' needs for information will continue to grow as demands for healthcare rise in the face of diminishing funding. Notwithstanding this financial pressure, the use of outcome data for clinicians can be illustrative for their own development, and ultimately and most importantly for maintaining high standards of management for those patients in their care. The electronic data capture system has been shown to be feasible while using a choice of questionnaires. The systematic review of the Bournemouth Questionnaire (BQ) is a new addition to the literature, and it identified the value of the BQ for inclusion in the app. Although the implementation process is at an early stage it has identified suitable approaches to introducing new innovations into day-to-day osteopathic practice.

Appendices

Appendix 2.1: Participant Information Sheet

Information sheet

A qualitative study to identify patients' views on the use of Patient Reported Outcome Measures in clinical practice: information for participants

We would like to invite you to be part of this research project. We are looking for patients to take part in an interview to help us understand the good things or challenges about using patient reported outcome measures (PROMS) in osteopathy. A PROM is a type of questionnaire, for example it could ask you about different things associated with your symptoms, or how you are feeling as a result of your treatment.

The study involves taking part in a short telephone interview. We would like to learn if you have filled in a PROM-questionnaire before, whether you feel you would be happy to use them in the future, and whether or not you feel they are helpful when you come for treatment. It will last about half an hour. The interview will be recorded. This lets us type up what you have said. It will be done without identifying you, and if we use any quotes in publications, this will be done also without identifying you. Once the audio recording has been typed up and checked to make sure it is accurate, it will be destroyed. You should only agree to take part if you want to; it is entirely up to you. If you choose not to take part this will not affect your treatment, and you will hear no more about it. Please read the following information carefully before you decide to take part. This will tell you why the research is being done, and what you will be asked to do if you take part. Please contact us if there is anything that is not clear, or if you would like more information. If you decide to take

part you will be asked to sign a consent form to say that you agree. You are still free to change your mind at any time and without giving a reason.

If you have any questions about the study, please contact the researcher, Carol Fawkes either by telephone (07732178308) or email (c.fawkes@qmul.ac.uk).

This study has been reviewed and approved by the Queen Mary Ethics of Research Committee (QMREC2013/57), Room W117, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Appendix 2.2: Participant Consent Form

Consent form (Version 2)

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.
Title of Study: A qualitative study to identify patients' views on the use of Patient Reported Outcome Measures* in clinical practice.
Queen Mary Ethics of Research Committee Ref: QMREC2013/57Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part.

- If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.
- *I understand that if I decide at any other time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately.*
- *I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.*

Participant's Statement: I _____ agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and below and the Information Sheet about the project, and understand what the research study involves.

Signed:
Date:

Investigator's Statement: I, Carol Fawkes, confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the volunteer.

**** A Patient Reported Outcome Measure (PROM) is a type of questionnaire that can ask you about different things associated with your symptoms, or how you are feeling as a result of your treatment.***

Appendix 2.3

TOPIC GUIDE – PATIENT INTERVIEWS

I would like to start off just getting your general views on the use of questionnaires/PROMs and then we can look at the specific PROMs I have sent to you.

Could you tell me first of all what you think about the general principle of completing PROMs about osteopathic treatment? Do you think that is a good thing or not?

What do you feel is a reasonable time to spend on these activities before they become burdensome;

Where do you think it would be better to complete something like that? Would you prefer to do it at home or go to the practice a little earlier and do it there?

Would you feel at all uncomfortable being asked to complete a PROM at the practice?

What is feasible for you to do when contributing data in terms of how you contribute information e.g. on paper, online, or by using a mobile device?

What type of support do you think would be necessary to encourage you to use a particular type of format for data collection *e.g.* using paper, online, or a mobile device?

Do you have any concerns about routine data collection and using PROMs?

Do you have any concerns about your data being sent to a third party e.g. a university research department?

What sort of information from this data collection would it be useful for you to know?

How would you like to see data from the questionnaires to be presented to make this information useful to you?

What sort of format would it be better to see this type of information e.g. would it be better to see this as a poster in the practice, as a leaflet to take away, or as an email attachment.

Look at individual PROMs.

A (BQ). What are your general impressions of this questionnaire? Is the numerical scale OK to use?
Do you feel the questionnaire measures everything that is relevant to you?

B (RMDQ)What are your general impressions of this questionnaire?
Do you feel the questionnaire measures everything that is relevant to you?

C (ODI)What are your general impressions of this questionnaire?
Do you feel the questionnaire measures everything that is relevant to you?

What other information do you feel it would be useful to collect in addition to what is contained in the PROMs we have seen earlier?

Is there anything else you feel we haven't covered that you would like to add?

Just to finish off would you mind if I asked which age band you fall into:
18-30
31-40
41-50
51-60
61-70
71 -80
81-90
91 or over

Would you like to be kept informed about how the project is developing with a copy of the project newsletter?

If yes by post or email.

Appendix 3.1: **Participant Information sheet – focus groups**

A qualitative study to identify osteopaths' views on the use of Patient Reported Outcome Measures in clinical practice: information for participants

We would like to invite you to be part of a research project, we are looking for osteopaths to take part in a focus group, to help us understand the advantages and challenges surrounding the use of patient reported outcome measures (PROMS) in osteopathy. We would like to hear your views on the use on PROMs in everyday clinical practice when you are assessing the change in a patient's progress. We would like to learn if you have had previous experiences of using PROMs or whether you would like to use them in the future, or whether you do not feel that they are helpful for your practice and patients.

The study involves taking part in a group discussion, or focus group. There will be small number (5-7) of other osteopaths taking part also. The focus group will be held in a convenient location and last about one hour. The focus group will be audio recorded and transcribed verbatim. All identifiable information will be removed from the transcript and any quotes used for publication will be anonymised. Once the audio recording has been transcribed and checked it will be deleted. You should only agree to take part if you want to; it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it. Choosing not to take part will not affect your access to future PROM services in any way.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please contact us if there is anything that is not clear, or if you would like more information. If you decide to take part you please sign the attached form to say that you agree.

You are still free to withdraw at any time and without giving a reason.

If you have any questions or concerns about the manner in which the study was conducted please, in the first instance, contact the researcher responsible for the study. If this is unsuccessful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Appendix 3.2

Consent form

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: A qualitative study to identify osteopaths' views on the use of Patient Reported Outcome Measures in clinical practice.

Queen Mary Ethics of Research Committee Ref: QMREC1207

Thank you for considering taking part in this research. The person organizing the research must explain the project to you before you agree to take part.

- If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.
- *I understand that if I decide at any other time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately.*
- *I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.*

Participant's Statement: I _____ agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and the Information Sheet about the project, and understand what the research study involves.

Signed:

Date:

Investigator's Statement: I Carol Fawkes confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the volunteer.

Appendix 3.3: Topic Guide Questions

Do you know what an outcome measure is?

Do you currently use outcome measures in clinical practice?

Have you thought about using outcome measures in clinical practice?

What information do you feel it would be useful to collect in addition to what is contained in the PROMs we have seen earlier?

Do you plan to use outcome measures to support your practice in the future?

What barriers, if any, do you perceive in using outcome measures as part of your day-to-day practice?

What is feasible for you to do when contributing data in terms of how you contribute information *e.g.* on paper, online, or by using a mobile device?

What do you feel is a reasonable time to spend on these activities before they become burdensome?

What type of support do you think would be necessary to encourage you to use a particular type of format for data collection e.g. using paper, online, or a mobile device?

What fears, if any, do you have about routine data collection and using PROMs?

What support, if any, do you feel you might need to participate in collecting data and using outcome measures?

How would you like to see data from the questionnaires to be presented to make this information useful to you?

Do you feel you might need any additional educational input to interpret the presented data?

Appendix 3.4: Initiatives to recruit physiotherapy interview participants.

Date	Contact	Response
03-07-2014	Contact with Chairperson of PPEF	Request for further information about project which was sent.
06-09-2014	Reminder email to Chair of PPEF	No response
10-11-2014	Reminder email to Chair of PPEF by third party	No response
18-01-2015	Email to QMUL physiotherapy alumni	No response
19-01-2015	Contact with Plymouth University physiotherapy alumni via third party	No response
03-02-2015	Forty personal letters sent to physiotherapists (selected from PhysioFirst database) in Leicester, Oxford, Cardiff, Manchester, Glasgow, Eastbourne, and Nottingham	1 response (Glasgow) declined to take part;1 response (Eastbourne) stating now retired.1 volunteer (Manchester) – interviewed 10-03-2015
05-02-2015	Letters sent to three physiotherapists also qualified as osteopaths	No response
06-02-02105	Contact with MACP research officer	Forwarded information about MACP (after reminder email) 20-03-2015
17-02-2015	Reminder to Plymouth University physiotherapy alumni via third party	No response
19-02-2015	Email forwarded to hub members to forward on to their physiotherapy colleagues in multidisciplinary practices	No response
05-03-2015	Contact with physiotherapists at Warwick Medical School	1 research physiotherapist (interviewed 24-03-

		2015);I research and physiotherapist clinician (interviewed 15-05-2015)
18-03-2015	Advert published in *Frontline* magazine	No response
23-03-2015	Application submitted to access MACP members	MACP finally tweeted message 07-05-2015
12-05-2015	MACP research officer	Interview held

Appendix 4.1: Search strings for literature search

Medline (via Ovid)(Roland Morris OR Roland Morris Disability Questionnaire OR Roland Morris Questionnaire OR Roland Disability Questionnaire OR Oswestry Disability Index OR Oswestry Index OR Oswestry Disability Questionnaire OR Bournemouth Questionnaire) AND hasabstract AND (instrumentation[sh] OR methods[sh] OR Validation Studies[pt] OR Comparative Study[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR outcome assessment[tiab] OR outcome measure*[tw] OR "observer variation"[MeSH] OR observer variation[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab]

OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])) NOT (("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) OR ("animals"[MeSH Terms] NOT "humans"[MeSH Terms]) OR (cancer[sb] OR veterinary[sb] OR aids[sb] OR bioethics[sb] OR jsubsetd OR jsubsets OR jsubsete OR jsubsetq OR jsubsetqis) OR "Arthritis, Rheumatoid"[Mesh] OR ("Nervous System Diseases"[Mesh] NOT ("Headache Disorders"[Mesh] OR "Neurologic Manifestations"[Mesh] OR "Neuromuscular Diseases"[Mesh])))

EMBASE((Roland Morris OR Roland Morris Disability Questionnaire OR Roland Morris Questionnaire OR Roland Disability Questionnaire OR Oswestry Disability Index OR Oswestry Index OR Oswestry Disability Questionnaire OR Bournemouth Questionnaire) AND ('questionnaire'/exp OR 'named inventories, questionnaires and

rating scales'/exp OR ‘psychometry’/exp OR ‘outcome assessment’/exp OR 'pain assessment'/exp OR 'disability'/exp OR 'validity'/exp OR 'reliability'/exp) AND [english]/lim AND [humans]/lim AND [embase]/lim AND ([article]/lim OR [review]/lim) AND [abstracts]/lim AND [1974-2009]/py) NOT (‘neoplasm’/exp OR ('neurologic disease'/exp NOT ('spinal cord disease'/exp OR 'headache and facial pain'/exp AND 'neuralgia'/exp AND 'radiculopathy'/exp)))

PsycInfo(exp (Roland Morris OR Roland Morris Disability Questionnaire OR Roland Morris Questionnaire OR Roland Disability Questionnaire OR Oswestry Disability Index OR Oswestry Index OR Oswestry Disability Questionnaire OR Bournemouth Questionnaire)/ OR exp spinal column/) AND (exp measurement/ OR exp test construction/ OR exp interrater reliability/ OR exp statistical analysis/)

Appendix 4.2 PRISMA 2009 checklist

Section/ topic		Checklist item	Reported on page
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	

Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I2) for each meta-analysis.	
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of	21	Present results of each meta-analysis done,	

results		including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097. From http://prisma-statement.org/documents/PRISMA%202009%20checklist.pdf.

Appendix 5.1. Participant Information Sheet for patients in the test-retest strand of the pilot

Information sheet (Version 2)

Research study: Pilot study to investigate the feasibility of collecting patient reported outcome measurement (PROM) data in osteopathic practice using a web app or a mobile phone app**: information for participants**

We would like to invite you to be part of this research project, if you would like to. You should only agree to take part if you want to; it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it. Choosing not to take part will not affect your access to treatment or services in any way.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please ask if there is anything that is not clear or if you would like more information. If you decide to take part your completion of the questionnaire will be taken that you have given your consent to participate.

You are still free to withdraw at any time and without giving a reason.

What does the project involve?

This project is looking at how easy it is to collect information about how you feel after treatment. We are using the Internet or a mobile phone app.

What will I have to do?

We will ask you if you would like to fill in a questionnaire using one of these ways. This will happen once before you have treatment.

We will then ask you if you would like to fill in a second questionnaire. This will happen one week after treatment. We would like you to accept a voucher for £5 which you be able to receive your after you have filled in the final questionnaire. You will need to contact the researcher, Carol Fawkes, using the voucher form to claim your voucher. This form will be available on the NCOR website, and is attached to this information sheet.

Why are we doing this project?

We are doing this project to try and find out how you feel after treatment. This information will be looked at by researchers from the National Council for Osteopathic Research (NCOR). The information we collect will not include your name, address, or date of birth.

We hope the study will help to improve osteopathic treatment. Your information will not be made available to anyone else. We will not sell it or use it for any commercial purposes. If you complete the questionnaire, we accept you have consented to take part in this study. It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep.

There is some more information available on a separate sheet about the different systems used in mobile phones. This is in case you think you might like to fill in the questionnaire using your mobile phone but are not sure which type of phone you have. It is important for us to make you aware that there is an extremely low risk that your phone could be hacked. If this happened the information you provide in the questionnaire could be viewed. This could happen if your phone is stolen, for example.

If you have any questions or concerns about how this study was carried out, you should firstly contact the researcher responsible for the study, Carol Fawkes (email: c.fawkes@qmul.ac.uk, or telephone 07732178308). If this is not successful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Appendix 5.2. Participant Information Sheet for patients in the responsiveness strand of the pilot

Information sheet

Research study: Pilot study to investigate the feasibility of collecting patient reported outcome measurement (PROM) data in osteopathic practice using a web app or a mobile phone app**: information for participants**

We would like to invite you to be part of this research project, if you would like to. You should only agree to take part if you want to; it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it. Choosing not to take part will not affect your access to treatment or services in any way.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please ask if there is anything that is not clear or if you would like more information. If you decide to take part your completion of the questionnaire will be taken that you have given your consent to participate.

You are still free to withdraw at any time and without giving a reason.

What does the project involve?

This project is looking at how easy it is to collect information about how you feel After treatment. We are using the Internet, or a mobile phone app.

What will I have to do?

We will ask you if you would like to fill in a questionnaire using one of these ways. This will happen once before you have treatment.

We will then ask you if you would like to fill two more questionnaires. This will happen one week after treatment, and six weeks after treatment. We would like you to accept a voucher for £5 which you will be able to receive after you have filled in the final questionnaire. You will need to contact the researcher, Carol Fawkes, using the voucher form to claim your voucher. This form will be available on the NCOR website and is attached to this information sheet.

Why are we doing this project?

We are doing this project to try and find out how you feel after treatment. This information will be looked at by researchers from the National Council for Osteopathic Research (NCOR)*. The information we collect will not include your name, address, or date of birth.

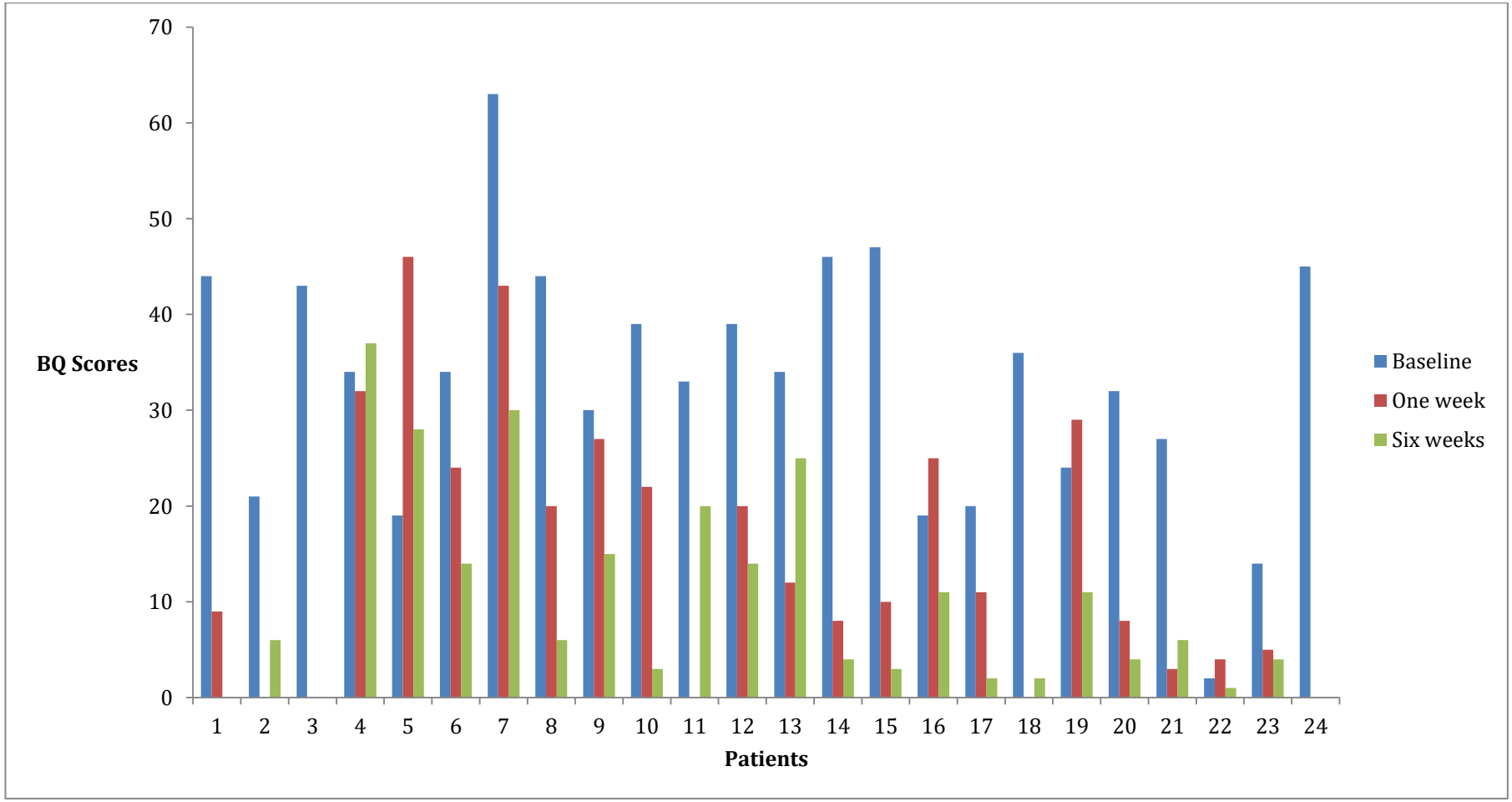
We hope the study will help to improve osteopathic treatment. Your information will not be made available to anyone else. We will not sell it or use it for any commercial purposes. If you complete the questionnaire, we accept you have consented to take part in this study. It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep.

There is some more information in a separate sheet about the different systems used in mobile phones. This is in case you think you might like to fill in the questionnaire using your mobile phone but are not sure which type of phone you have. It is important for us to make you aware that there is an extremely low risk that your phone could be hacked. If this happened the information you provide in the questionnaire could be viewed. This could happen if your phone is stolen, for example.

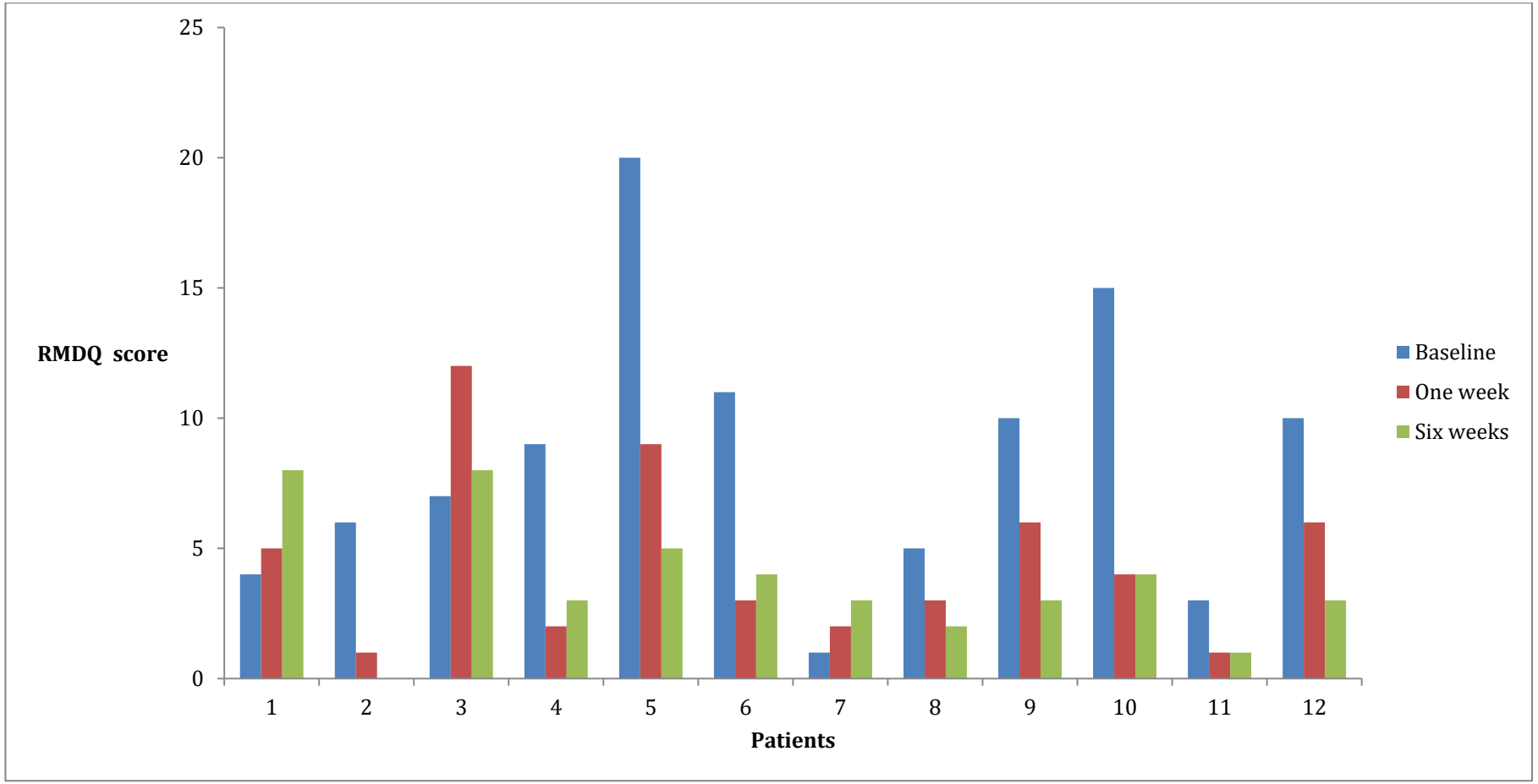
If you have any questions or concerns about how this study was carried out, you should firstly contact the researcher responsible for the study, Carol Fawkes (email: c.fawkes@qmul.ac.uk, or telephone 07732178308). If this is not successful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

* Carol Fawkes
National Council for Osteopathic Research, Barts and The London School of Medicine and Dentistry, Centre for Primary Care and Public Health, Blizard Institute,
Yvonne Carter Building, 58 Turner Street, London, E1 2AB

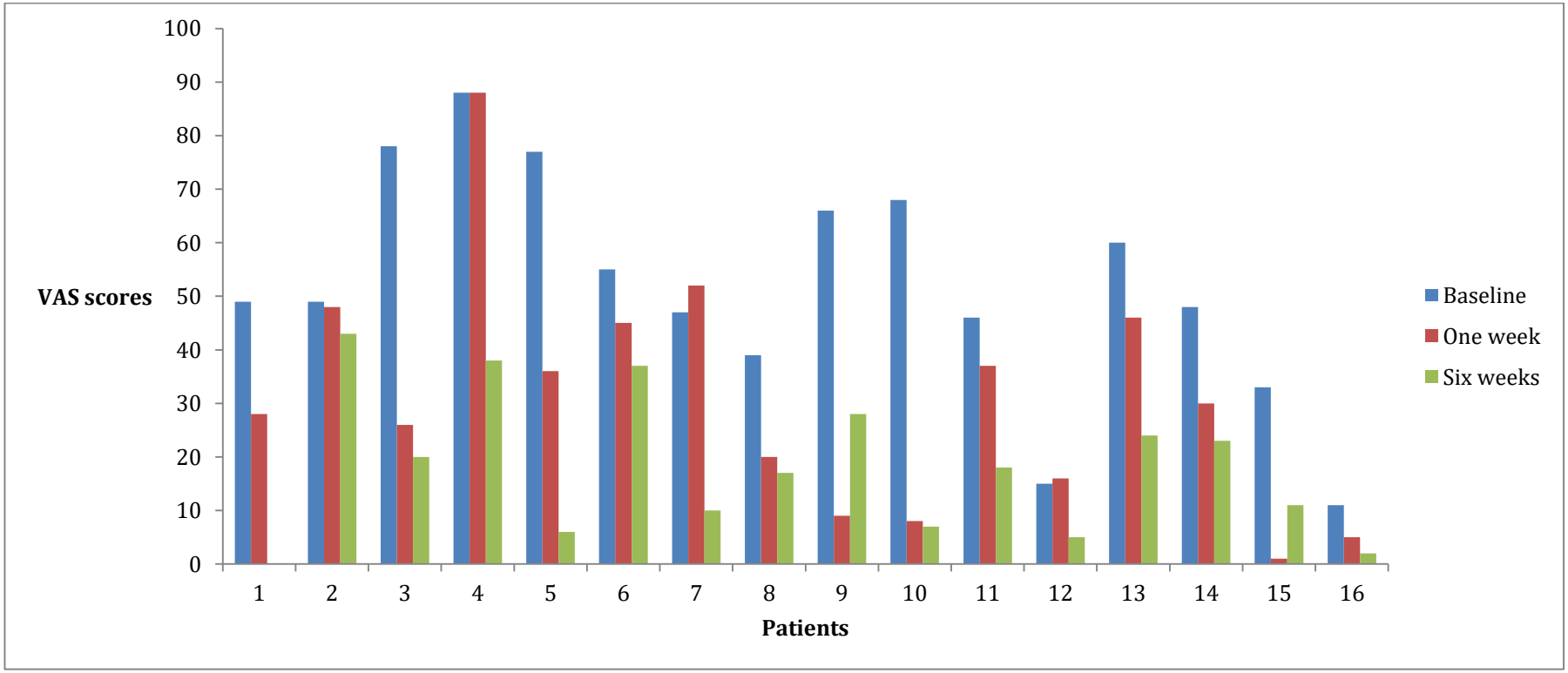
Appendix 5.3. Change in BQ score for all patients at each time point in the responsiveness strand of the pilot



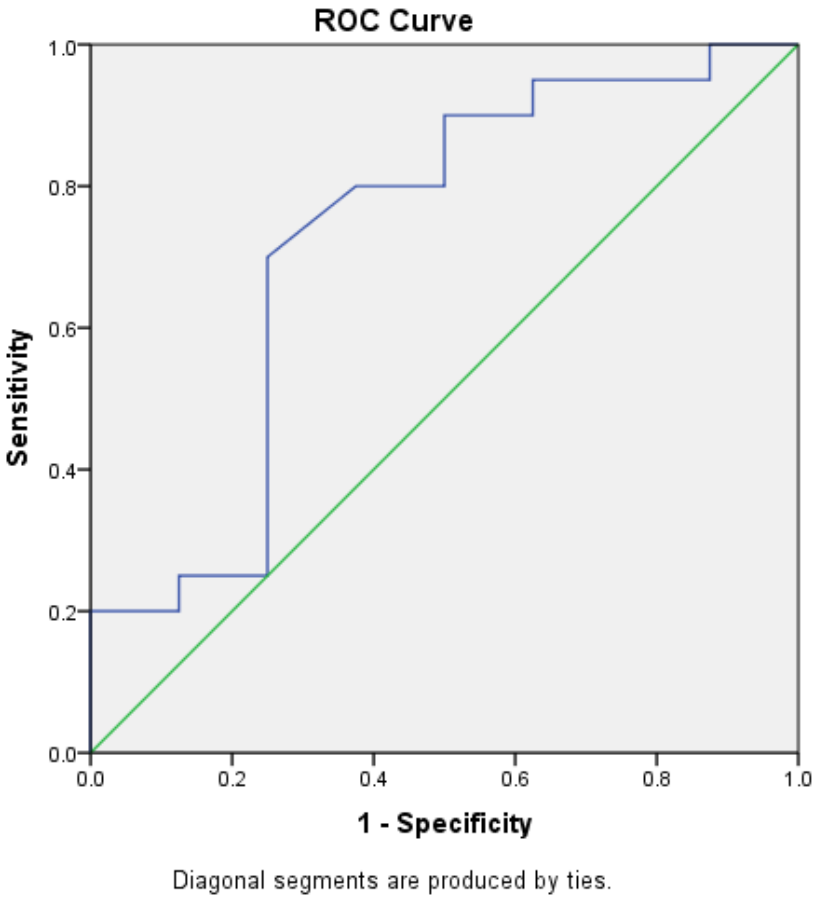
Appendix 5.4. Change in RMDQ score for all patients at each time point in the responsiveness strand of the pilot



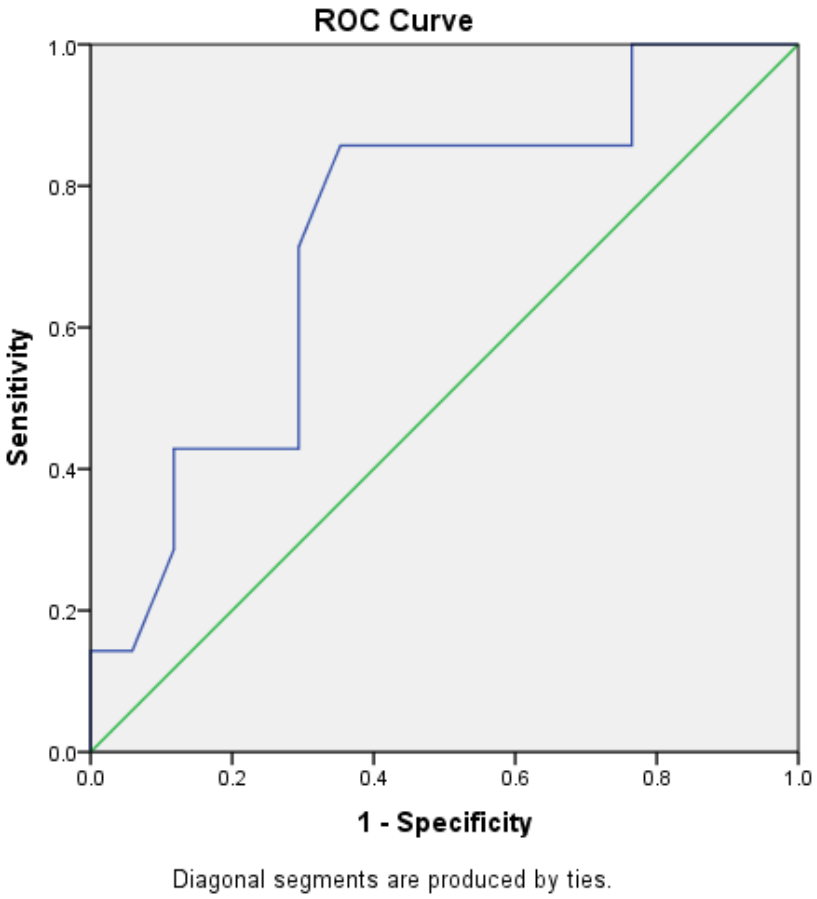
Appendix 5.5. Change in VAS score for all patients at each time point in the responsiveness strand of the pilot



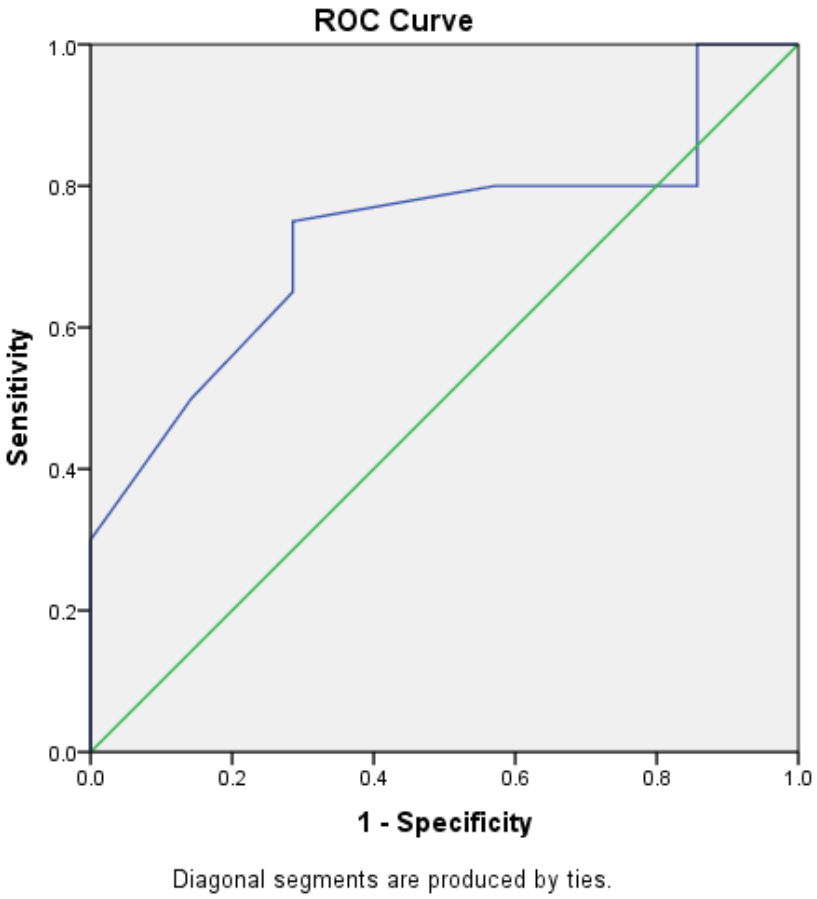
Appendix 5.6 **Receiver Operator Characteristic (ROC) curves for the BQ, VAS, and RMDQ**



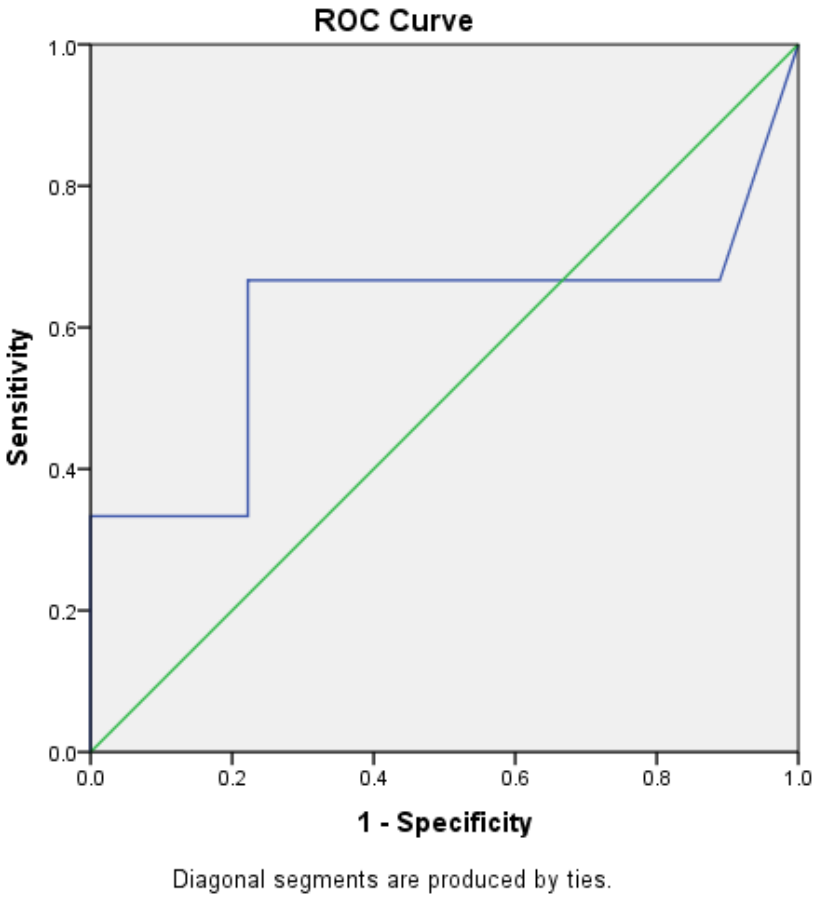
BQ 1 week post treatment


ROC Curve
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Diagonal segments are produced by ties.

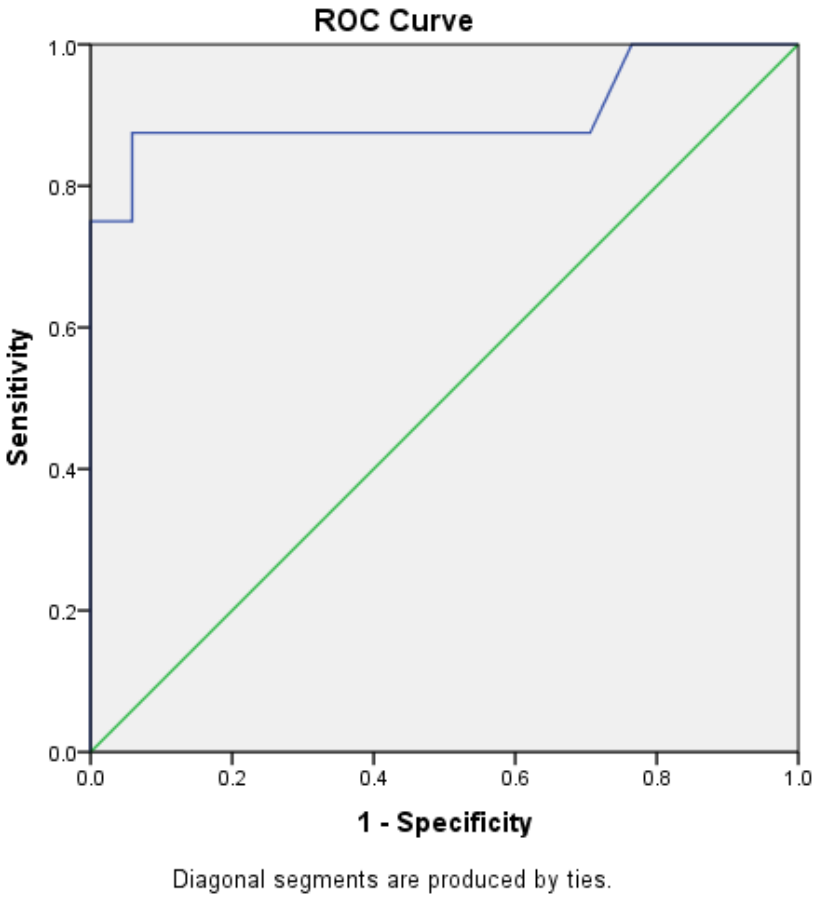
BQ 6 weeks



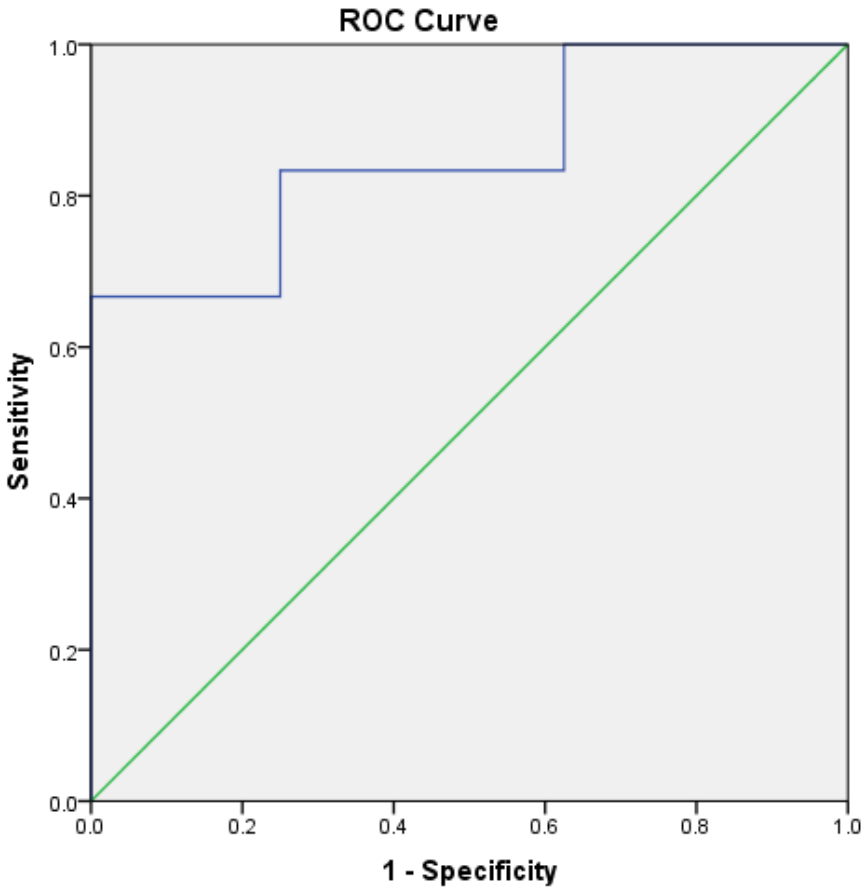
RM score 1 week post-treatment



RM 6 weeks post-treatment



VAS at 1 week post-treatment



VAS at 6 weeks post-treatment

Appendix 6.1: Participant information card

National Council for NCOR Osteopathic Research

Thank you for using the NCOR data collection facility. All of your data is kept securely and anonymously. It will be used for research purposes only. It will not be sold to any commercial organisations.

What do I need to do?

There are three options depending on whether you are using a desktop or laptop computer, an Android phone or Tablet, or an iOS (Apple) phone or Tablet. Firstly, please go to the website http://www.clinvivo.com/ncor/

If you are using the Internet on a desktop or laptop or an Ipad: please click on the computer image.

If you are using an Android mobile or tablet to download our app, please click on the image of the green android figure as shown below.

Instructions about how to download the app are available on a separate sheet – please ask your osteopath for this.

If you prefer, you can use the QR code

If you are using an iOS (Apple) mobile or tablet to download our app, please go to the App store, type in Clinvivo, and click install. There is no charge for this app. If you prefer you can use the QR code for the iOS version:

Contact information

If you have any queries concerning the project or using the website or mobile app, please contact the project lead, Carol Fawkes. Email: c.fawkes@qmul.ac.uk or telephone:07494059509.

Appendix 6.2 **Patient newsletter**

Patient newsletter

December, 2015

Queen Mary University of London

Dear Patient,

Last year you were kind enough to spend time participating in an interview. At the time we were in the process of developing a system for the collection of data about outcomes of osteopathic treatment.

In total, 22 patients agreed to be interviewed for this project as we tried to find out your views about collecting data in osteopathic practice, and what you thought about three questionnaires. We also interviewed clinicians (osteopaths, physiotherapists, and chiropractors) as part of the project to get their views concerning the collection of data about outcomes of care.

In summary, you told us:

- You were broadly in favour of the idea of practice-based data collection;
- You felt that collecting data would provide evidence for the profession to develop;
- You were happy to complete questionnaires either at home or in the practice depending on what was the most convenient for you;
- Many of you were happy with the idea of using the Internet, a tablet device, or a smartphone app;
- Some of you mentioned you did not use the Internet, and were concerned that patients might be excluded if paper data collection was not available;
- There were differing opinions on the questionnaires (Patient Reported Outcome Measures or PROMs) discussed but you felt it was important to include options which had words alone as well as numerical scales;

- You mentioned that any information available about practice-based data collection should be quite brief and specific. You stated that you would not want your data shared outside of the research team, and it should not be made available to any commercial organisation;
- You mentioned that there were key things associated with your symptoms that the PROMs did not collect;
- You did not want PROM data collection to interfere with the consultation process.

What is happening with the project now?

In July, 2014 we started to pilot a data collection system using the Internet (a web app) or an app for a smartphone or Tablet. This pilot process took place in the training colleges (Osteopathic Educational Institutions) and in private practices throughout the UK.

The pilot process focussed solely on spinal pain. This has helped us to focus on how well the system worked, and how well the questionnaires were completed by patient volunteers. The system has performed very well. It allows information to be collected at the first appointment, one week after treatment, and then one final questionnaire at six weeks after treatment. Questionnaires have been well

completed, although some of the later questionnaires (at 6 weeks after treatment) have tended to be forgotten. The data we have collected have allowed us to look at a range of different items about practice. This includes from the initial appointment:

- Knowing more about the patients attending for treatment (e.g. age, sex, work status, and ethnic background);
- The main reason for seeking treatment (e.g. pain, stiffness, advice);The main area of symptoms experienced;
- How long symptoms have been present;
- The severity of symptoms initially (from completion of a specific questionnaire).Data collected at the follow up appointments at one week and six weeks after treatment include:

- The change in symptoms since treatment began (from completion of a specific questionnaire);
- How satisfied you have been with treatment;
- Your experience of osteopathic care;
- Your overall change in symptoms since treatment began.

All of the data collected using the data collected system are anonymous, and will not be shared with any third parties.

What are the next steps for the project?

Now that the system is functioning well, we are encouraging osteopaths across the country to use the data collection system in their practices. We will continue to develop the system to add questionnaires for different areas of the body e.g. shoulders, knees *etc,* and for different symptoms. There has been considerable interest from other healthcare professions in the work we have undertaken. We have presented the findings of this study at national conferences (Egham, Bournemouth, and Nottingham), and international conferences (Sao Paulo, Montreal, and Rome).

Without your participation in this project, none of this work would have been possible. We are very grateful to you for taking the time to be involved. If you have any further questions about this project, please contact me either by email (c.fawkes@qmul.ac.uk), or by telephone (07732178308).

Appendix 6.3. Clinicians' newsletter

Research newsletter

December, 2015

Queen Mary University of London

Dear Colleague,

Last year you were kind enough to spend time participating in an interview. At the time we were in the process of developing a system for the collection of data about outcomes of osteopathic treatment.

In total, 22 patients agreed to be interviewed for this project as we tried to find out their views about collecting data in osteopathic practice, and what they thought about three questionnaires. We also interviewed clinicians (osteopaths, physiotherapists, and chiropractors) as part of the project to get their views concerning the collection of data about outcomes of care.

In summary, patients told us:

- They were broadly in favour of the idea of practice-based data collection;
- They were happy to complete questionnaires either at home or in the practice depending on what was the most convenient for them;
- Many of them were happy with the idea of using the Internet, a tablet device, or a smartphone app;
- They mentioned that any information available about practice-based data collection should be quite brief and specific. They stated that they would not want their data shared outside of the research team, and it should not be made available to any commercial organisation;

In summary, clinicians told us:

- They felt asking patients about their outcomes of care was a positive thing, and an increasingly accepted part of clinical practice;

- There were some practical issues to overcome associated with how the patient should be asked to complete the PROMs, and the availability of different formats for patients in different settings, and with different language capabilities;
- The data have a range of uses including personal reflection on practice, for evaluating patient care, and demonstrating results of practice to external agencies e.g. third party payers.

What is happening with the project now?

In July, 2014 we started to pilot a data collection system using the Internet (a web app) or an app for a smartphone or Tablet. This pilot process took place in the training colleges (Osteopathic Educational Institutions) and in private practices throughout the UK. The pilot process focussed solely on spinal pain. This has helped us to focus on how well the system worked, how well the questionnaires were completed by patient volunteers, and how well the questionnaires performed in an electronic format.

The system has performed very well. It allows information to be collected at the first appointment, one week after treatment, and then one final questionnaire at six weeks after treatment. The data we have collected have allowed us to look at a range of different items about practice. This includes from the initial appointment:

- Knowing more about the patients attending for treatment (e.g. age, sex, work status, and ethnic background);
- The main reason for seeking treatment (e.g. pain, stiffness, advice);
- The main areas of symptoms experienced;
- How long symptoms have been present;
- The severity of symptoms initially (from completion of a specific patient reported outcome measure (PROM).Data collected at the follow up appointments at one week and six weeks after treatment include:

- The change in symptoms since treatment began (from re-completion of the PROM used at the first appointment);How satisfied patients have been with treatment;
- Patients' experience of osteopathic care;
- Patients' overall change in symptoms since treatment began.
- All of the data collected using the data collected system are anonymous, and will not be shared with any third parties.

What are the next steps for the project?

Now that the system is functioning well, we are encouraging osteopaths across the country to use the data collection system in their practices. The modified system involves osteopaths having a code which is linked to their GOsC number acting as a unique identifier for individual clinicians and their practices.

If you would like to sign up to use the revised PROM system in your practice, please contact me and I will arrange for codes to be created for you. We will continue to develop the system in the future to add questionnaires for different areas of the body e.g. shoulders, knees etc, and for different symptoms to accommodate patients who receive treatment for both musculoskeletal and non-musculoskeletal symptoms.

There has been considerable interest from other healthcare professions in the work we have undertaken. We have presented the findings of this study at national conferences (Egham, Bournemouth, and Nottingham), and international conferences (Sao Paulo, Montreal, and Rome). Publications from this study are currently being prepared for submission to peer-reviewed journals. An interim summary of the findings will be available in the osteopathic press early next year.

Without your participation in this project, none of this work would have been possible. We are very grateful to you for taking the time to be involved.

Contact details

If you have any further questions about this project, please contact me either by email (c.fawkes@qmul.ac.uk), or by telephone (07732178308).

Appendix 6.4: Data reported back to osteopaths

The data included:

- Baseline sociodemographic data including patients' age, sex, work status, and ethnic background;
- Service data *e.g.* waiting times for treatment, duration of symptoms, patients' main reason for seeking treatment;
- Symptom data including main area of symptoms, and the number of symptom areas reported;
- Ratings of satisfaction;
- Ratings of experience;
- Global change since osteopathic treatment began.

Analysis of PROM data to include:

- The raw scores at pre- and 6 week post-treatment stages (Little and MacDonald, 1994; Farrar *et al.,* 2001; Hurst and Bolton, 2004; Gurden *et al*., 2012);
- Percentage change score (Little and MacDonald, 1994; Farrar *et al*., 2001; Hurst and Bolton, 2004; Gurden *et al*., 2012);

Appendix 7.1: Outcome data collected in paediatric studies involving manual therapy listed on PubMed.

Symptoms/disorders	Outcome measured	Manner in which outcome measured
Infantile colic	Denckens *et al.,* 1996 Information not available	Information not available
	Wiberg *et al.,* 1999 Periods of sleeping Periods of being awake Periods of crying Bowel movements Feeding patterns	Daily diary
	Mercer *et al.,* 1999 Patients' perception of infant's response to treatment Presence/absence of colic	5-point Likert scale Completely recovered (1) Somewhat better (2) The same (3) Somewhat worse (4) Much worse (5)
	Olafsdottir *et al.,* 2001 Hours of crying per day from baseline to each of three visits Global change in symptoms Presence/absence of colic	Crying diary Global improvement scale5-point Likert scale Completely well (5) Marked improvement (4) Some improvement (3) No improvement (2) Getting worse (1)
	Koonin *et al.,* 2002 Duration of crying per day Frequency of crying per day	Crying diary Measured on ordinal scale as reported by parental

	Total crying per day Improvement	questionnaires pre-treatment, post-treatment, and at follow-up.
	Gladovatz *et al.,* 2003Case study	Pre-and post-treatment questionnaire
	Heber *et al.,* 2003 Daily hours of crying between week 1 and week 5 Intensity of crying	Daily crying diary Numerical rating scale
	Karpelowski *et al,* 2004 Duration of crying Frequency of crying Total amount of crying	Questionnaire
	Hayden and Mullinger, 2006 Daily hours of crying between week 1 and week 4 Daily hours of sleeping between week 1 and week 4 Daily hours of infant being held or rocked (taken as an indication of low-level colic)	Daily crying diary
	Davies *et al.,* 2007 Information not available	
	Browning *et al.,* 2008 Amount of daily crying Duration of daily crying Typical time of day when colic behaviour occurred Amount of sleep Amount of non-distressed awake behaviour.	Daily crying diary

	Mills *et al.,* 2010 Retrospective case-matched study	 Information not available
	Miller *et al.,* 2010 Reduction in crying to <2 hours per day >30% improvement in daily crying hours Presence/absence of colic	
	Wiberg *et al.,* 2010 Temper tantrums Falling to sleep (sic) Sleep through night	 Rare/never 1-2 daily >3 daily <20 minutes >20 minutes Yes No
	Periods of sleeping (Miller *et al.,* 2012) Periods of being awake Periods of crying Bowel movements Feeding patterns Global change in symptoms Adverse events	 24 hour diary Global improvement scale Parental report
	Miller and Phillips, 2009. Survey of children (n=117) receiving or not receiving past treatment for colic. Behaviour (number of tantrums) Falling asleep	 Rare/never 1-2 daily >3 per day <20 minutes

	Staying asleep	>20 minutes Yes No
Nocturnal enuresis	Blomerth *et al.,* 1994 Spinal dysfunction Episodes of bed-wetting	 Manual examination Case notes recording
	Bolin 2010 Range of motion Restrictions in spinal movement	 Physical examination Physical examination
	Bosler 1979 Not disclosed	 Information not available
	Gemmell 1989 Not disclosed	 Not disclosed
	Leboeuf et al., 1991 Baseline measure of bed wetting Mean number of wet nights/week at end of first 2 weeks Adverse events	 Not disclosed Not disclosed Not disclosed
	Reed *et al.,* 1994 Number of wet nights per 2 weeks at follow up Success of treatment	 Not disclosed Defined as 50% improvement or greater
	Van Poecke and Cunliffe, 2009 – case series (n=33) Wet night frequency Number of treatments Presence of constipation Diurnal urinary output	 Baseline, 3,6,9, and 12 months Case records Case records Comparison to standard age charts
Asthma	Balon *et al.,* 1998	

	Spirometry Asthma symptoms Need for inhaled β-agonists Use of oral corticosteroids Overall satisfaction with treatment Adverse events	FEV_1Paediatric Quality of Life Questionnaire Symptoms (10 questions) Activities (5 questions) Emotions (8 questions) Number of puffs Diary Scale not disclosed Method not disclosed
	Bronfort *et al.,* 2001 Spirometry Fixed loop volume Lung volume Plethysmography Bronchial challenge with exercise Asthma severity Patient-rated quality of life Parent/guardian-rated quality of life β2-agonist use Wheezing Shortness of breath Coughing Disturbed sleep Feeling of panic Restricted activity Overall treatment satisfaction Overall treatment satisfaction	 Questionnaire with 11 item NRS Questionnaire Questionnaire Number of puffs per day Asthma diary Asthma diary Asthma diary Asthma diary Asthma diary Asthma diary Patient questionnaire Parent/guardian questionnaire

	Asher *et al.,* 1990 Pulmonary function Symptom recurrence Quality of Life Days in hospital	 RV; TLC; PEFR; FEF25 to 75; FEV1 measured before and after 20 min salbutamol inhalations, and immediately following the first and fourth treatments; best peak flow rate at discharge Admit/relapse No formal measure Not disclosed
	Field *et al.,* 1998 Pulm function tests Quality of Life Saliva cortisol levels Behaviour (affect, anxiety, activity, vocalizing) for 30 min before and after first and last treatments	 FVC, FEV1, and FEF25 to 75 at days 1 and 30; PEFR each night State Anxiety Scale parents and children videotaped behaviour of child
	Nikooee *et al.,* 2008 Pulmonary function tests (FEV, FVC)	 Spirometry
Hyperactivity	Gieson *et al.,* 1989 ANS system activity Motion/activity measurement during tasks Spinal biomechanics Spinal distortion Parental rating of activity	 Electrodermal evaluation Chiropractic evaluation X-Ray Not disclosed
Autism spectrum disorder	Hoffman and Russell, 2008 Social interaction Language skills Muscular activity Paraspinal muscular lesions	 Electromyogram Thermal scanning

	Aguilar *et al.,* 2005 Activities associate with Autism-Spectrum disorder Paraspinal muscular lesions	 Modified Autism Checklist Childhood Autism Rating Scale Pre- and post-treatment radiographic scan Brain stem-evoked potential recordings Thermal scanning Supine leg change analysis
	Khorshid *et al.,* 2006 Communication Verbal skills Ability to make eye contact Improved mood Physical sports skills	 Autism Treatment Evaluation Checklist (ATEC) Leg length analysis Pre-treatment and completion X-Rays
Growing pains	Alcantara and Davis, 2010 Night waking due to pain Segmental motion change	 Parental reporting Palpation
	Uziel *et al.,* 2010 Pain threshold Sleep quality Days off school with pain Location of pain Duration of pain Frequency of pain Development of other pain syndromes Use of analgesia measures (Including CAM)	 Fisher-type dolorimeter Parental and self-report: Good/Moderate/Poor
	Bowers 1997 Information not available	

	Eriksen 1996 Information not available	
Myasthenia Gravis	Alcantara *et al.,* 2003 Skin temperature analysis Segmental palpation Muscle hypertonia Range of movement Radiographic examination	 Dual probe hand-held skin surface instrument Manual therapy Manual therapy Manual therapy Full spine and weight-bearing
Otitis media	Degenhardt and Kuchera, 2006 Post-treatment recurrence of findings within 12 months Change in somatic dysfunction	 Clinician recording Osteopathic evaluation – resolved/Improved/ Unresolved
	Erickson 2006 Recurrence of otitis media episodes Palpatory evaluation Developmental maturity	 Not disclosed Profile of Development (PoD)
	Mills *et al.,* 2003 Frequency of AOM episodes Surgical interventions Behaviour change including irritability, disobedience, ear-pulling, clumsiness, listening when spoken to, restful sleep, and hearing when spoken to. Condition of the middle ear	 Recorded at study visits Recorded at study visits 5-point Likert scale (5= much more; 1= much less)

		Tympanometry (monthly)
	Sawyer *et al*., 1999 OM assessment Use of medication Use of medical services Sleeping quality (very well, slightly restless, extremely restless) Exhibiting symptoms of infection (including pulling/biting ear, fussiness/crying, irritability, complaint of ear pain, clinginess, fever, or runny nose).	 Tympanometry Otoscopy Parental diary Parental diary Parental diary
	Wahl *et al.,* 2008 OMT and Echinacea purpurea RCT Otitis media assessment Occurrence of AOM Number of episodes of AOM Side-effects of treatment Beliefs about treatment group assignment	 Otoscopy (baseline, 3,6, and 6 months. Telephone interview Telephone interview Telephone interview Telephone interview
Middle ear effusion	Steel *et al.,* 2014 Rate of vibration of tympanic membrane Ability of tympanic membrane to reflect sound Somatic dysfunction	 Tympanometric readings (weekly) Acoustic reflectometer readings (weekly) OMT palpatory evaluation

Complex Regional Pain Syndrome	Beck 2009 Case study Hypersensitivity to touch and pressure Dermographia	 Physical examination Physical examination
Scoliosis	Lantz and Chan 2001 Change in spinal curvature	 X-Ray
	Rowe et al., 2006 Cobb angle Quality of Life Expectation for improvement	 X-Ray Scoliosis Quality of Life Index (SQLI)5-point Likert scale: very much improved to very much worsened
	Lewit and Tesarova, 1961	Information not available
Sports injuries	Almeida 2011: Swimmer's shoulder problem Clinical evaluation and testing Range of Motion Pain Disability	 Palpation Clinical tests including Neer, Hawkins-Kennedy, anterior apprehension, Jerk test, and sulcus sign; Goniometer Vas-Pain Disabilities of Arm, Shoulder, and Hand (DASH) Questionnaire
	Woo, 1993: high jump injury Neurological status Muscle tone	 Reflexes Clinical evaluation
Kinetic Imbalance due to Suboccipital Strain (KISS)	Brand *et al.,* 2005 No clinical trials were found that evaluated	 None disclosed

Syndrome	the effects of manual therapy or osteopathy	
Attention Deficit Disorder (ADD)/Attention Hyperactivity Deficiency Disorder (ADHD)	Accorsi *et al.,* 2014 Changes in somatic dysfunction Neuro-behavioural change Adverse events	 Anatomical assessment Biancardi-Stroppa Modified Bell Cancellation Test Case note recording
	Alcantara and Davis, 2010 Number of visits Medication use Sites of segmental dysfunction ADHD/ADD symptom change Adverse events	 Anatomical examination ADHD Monitoring Questionnaire Case note recording
	Bastecki, 2004 Case report Tics Behavioural issues Range of Movement Lateral shift of head and thorax Reversal of cervical lordosis	 Physical examination Physical examination X-Ray (Atlas plane line angle)
	Gillespie 2009 Cognitive development evaluation Brain cycle measurement	 Development Assessment of Young Children test Not disclosed
	Muir, 2012 ADHD symptoms (including school performance, home performance.	 Parental report

	Acting out, and ability to follow instructions). Orthopaedic evaluation Spinal curvature Range of motion Adverse events	 Kemp, Maigne, Jackson, and Houle test) Chiropractic evaluation Chiropractic evaluation Chiropractic evaluation Clinical record keeping
Indigestion or heartburn	Bryner and Staerker, 1996 Symptom improvement Presence of mid back pain Indigestion relief	 Survey questionnaire (researchers' own design) Survey questionnaire Survey questionnaire
Pre-term/low birth weight babies	Cameron *et al.,* 2005 General assessment Posture Head lag Head control in sitting Infantile reflexes Spontaneous movement Muscle tone Birth weight Motor development	 Longitudinal Assessment of the Pre-term Infant (LAPI) Observation Observation Observation Observation Observation Observation Observation Alberta Infant Motor Scale (AIMS)
	Cerritelli *et al.,* 2013 Time in NICU Weight change General wellbeing	 Difference in days between entry and discharge Daily weight gain Number of episodes of vomiting Regurgitation Stooling Use of enema

	Economic evaluation	Time to full enteral feeding NICU costs
	Guzzetta *et al.,* 2011 Pre-term babies and massage Brain electrical activity	 EEG
Dyslexia	Bull, 2007 Parental stress Performance tests Wechsler Intelligence Scale for Children Coding Verbal tests Literacy tests Self esteem	 Parenting Stress Index Matrix analogies test Non-verbal reasoning Problem solving and deductive logic Draw a person WISC symbol search WISC digit span WISC vocabulary WISC information WISC similarities WISC comprehension Word literacy test Chapman and Turner reading self-concept Burden myself as learner scale
Infantile torticollis	Cheng *et al.,* 2001 Range of motion Head tilt	 Observation Observation
	Davis *et al.,* 2007 Asymmetry Range of motion Tissue texture abnormalities (oedema, muscle fibrosis) Motor function	 Observation Observation Physical examination Gross motor function scale

	Haugen *et al.,* 2011 Range of motion change by video analysis Symmetry of head position	 Worse No Significant Change Better Much Better Physiotherapy evaluation
	Hobaek Siegenthaler, 2015 Range of movement Head tilt	 Arthrodial protractor Clinician and parental evaluation
Congenital club foot	El-Hawary *et al* .,2008 Walking Club foot classification Rate of kinetic ankle motion	 Gait analysis Diméglio score Not disclosed
	Richards *et al.,* 2005 Severity of clubfoot Surgical intervention	 Dimeglio scale Moderate Severe Very severe Follow up data at 20-62 months
Temporomandibular joint (TMJ) disorders Limited mouth opening	Monaco *et al.,* 2008 Amplitude of maximal opening-closing movements Velocity of maximal opening-closing movements	 Kinesiographic recording
Back and neck pain	Dissing *et al.,* 2016 Overall change in symptoms Change in pain intensity after 2 weeks Satisfaction with treatment Total duration of complaint	 Global perceived effect 11-item NRS ND Time in weeks using weekly SMS

	Number of recurrences in symptoms during 3-27 months follow up period Average complaint time Pain site Quality of Life Expectations of treatment Expectation of future course (symptom return)	SMS Weeks using SMS Interview KID Screen 27 screening questionnaire Not disclosed Not disclosed
	Hayden *et al.,* 2003 Spinal subluxation Range of motion Pain severity Change in symptoms	 Clinical examination Clinical examination VAS 5-point Likert scale Worse Same Improved Much improved Resolved
Night waking	Dong et al 2009 Night waking	 Not Disclosed abstract in Chinese and no full text available
Cerebral palsy	Bennett 2007 Gait impairment Dynamic foot pressure Joint range of movement	 Observational Gait Scale Foot scan force plate Manual examination
	Duncan *et al.,* 2004 Parental perception and cerebral palsy Leg or hand use Sleep	 Qualitative interview

	Mood Speech or drooling Bowel movements Cognition Muscle stiffness Level of happiness	
	Hansen *et al.,* 2014 Gait change Gross motor abilities	 GAITRite system walkway Gross Motor Function Scale
	Wang *et al.,* 2008 Change in motor function Muscle tone	 Gross Motor Function Measurement -66 (GMFM-66) Modified Ashworth Scale
	Wyatt et al, 2011RCT (n=142) Motor function Quality of Life Behaviour rating Time to get to sleep Time asleep Sleeping change General health change Parent/carer quality of life Parental strength of belief in the benefit of cranial treatment	 Gross Motor Function Measure-66 (GMFM-66) Child Health Questionnaire (CHQ PF50) Paediatric Pain Profile (PPP) Sleep diary Sleep diary Global sleeping Global health SF-36 Questionnaire
Cyclic vomiting	Hubbard and Crisp, 2010 Abdominal pain Headache Cessation of symptoms	 Faces Pain Scale Faces Pain Scale Case notes

Headache	Weber-Hellstenius, 2009 Headache or neck pain prevalence Pain intensity Trigger point presence Cervical joint dysfunction Range of motion Head posture Muscle tone Frequency of headache or neck pain	 Questionnaire Likert scale Physical examination Physical examination Goniometer Physical examination Physical examination Questionnaire
Neurological deficits (including Duchenne or Becker Muscular Dystrophy	Nabukera *et al.,* 2012 Caregiver reports of use of massage and manipulation	 Not disclosed
Arthrogryposis	Thomsen *et al.,* 2010 Pain Walking ability	 Information not available Information not available
Gastro-oesophageal Reflux Disorder (GORD)	Jonasson and Knap, 2006 Nausea Neck pain Abdominal pain Vomiting after eating	 Parental diary Parental diary Parental diary Parental diary
Cancer	Field *et al.,*2001 Neutrophil recovery rates	 Information not available
	Post-White J, Hawks RG, 2004 Mood and anxiety	 Information not available
	Montgomery *et al.,* 2011 Health related quality of life	 SF-36
Respiratory disorders Infections	 Kline, 1965	

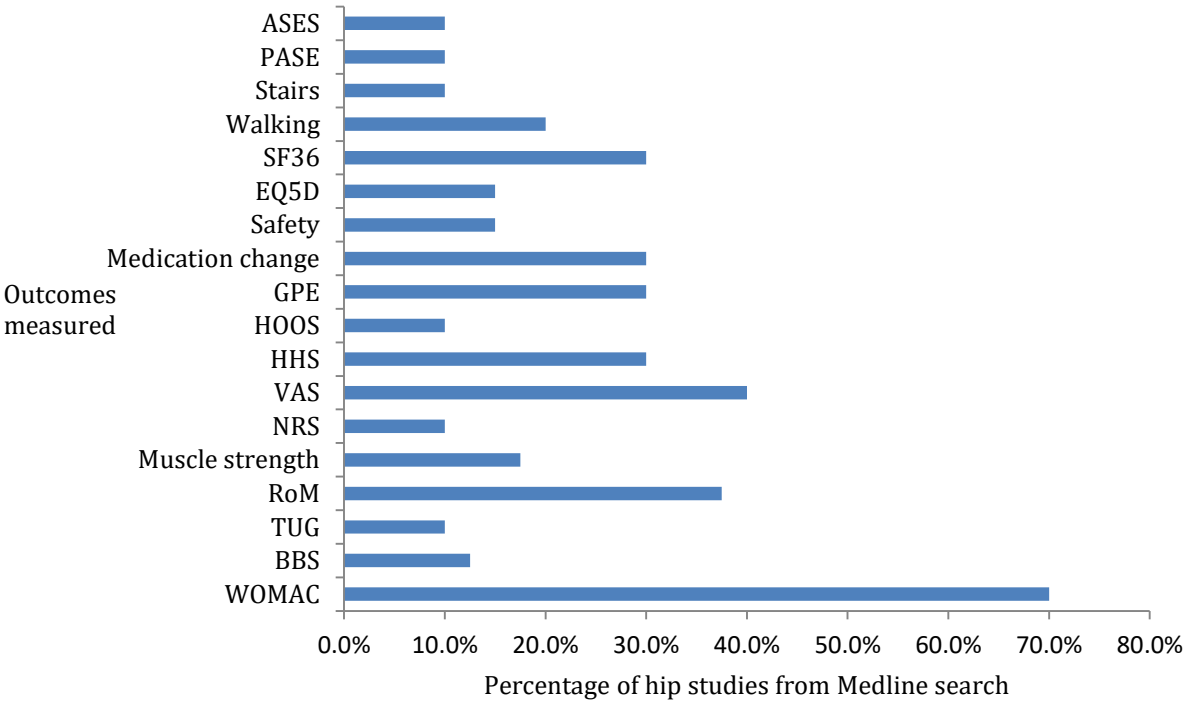
Pneumonia Bronchiolitis	Information not available Purse, 1966 Information not available Belcastro et al., 1984 Respiratory Frequency Presence or absence of wheezing, Intercostal retraction, Length of hospital stay, General health status	Not disclosed Information not available Not disclosed Not disclosed Not disclosed Hospital service data Radiograph, CBC, and microorganism culture
Head tilt	Knutson, 1996 Head tilt Spinal symmetry	 X-Ray
Cystic fibrosis	Hernandez-Reif *et al.,* 1999 Anxiety Mood Peak Expiratory Flow	 Not disclosed Not disclosed Not disclosed Spirometry
Chronic sleep bruxism	Knutson, 2003 Neck pain Headache Bruxism episodes	 VAS VAS Parental report
Gilles de la Tourette Syndrome	Kompoliti *et al.,* 2009 Massage and chiropractic treatment	 Not disclosed
Rett Syndrome	Lotan, 2007 Summary of case studies in different disciplines and approaches. Massage, chiropractic, myofascial release, and craniosacral therapy are discussed	 None disclosed

Suboptimal breast feeding	Miller *et al.,* 2009. Case series (n=114) Feeding at breast exclusively; Rating of improvement in of breastfeeding; Weight gain	 Mother's report Mother's report Mother's report
Joint hypermobility and benign hypotonia	Mintz-Itkin *et al*., 2009 Muscle tone Mobility Independent sitting, 4-point kneeling, Pull to standing, Onset of independent walking Gross Motor Development	 French Angles Factor of the Infant Beighton's mobility score Parental diary Parental diary Parental diary Parental diary Alberta Infant Motor Scale
Myopia	Neroev *et al.,* 2006 Uncorrected and subcorrected visual acuity, Ocular accommodation reserves, Myopiaregression rates Adverse events/ complications of treatment	 Not disclosed Not disclosed Not disclosed Not disclosed
Infantile postural asymmetry	Philippi *et al.,* 2006 Infantile position awake and asleep, time for carrying or 'kangarooing', time for putting the infant in a car seat. Vegetative parameter changes in vomiting, sleeping, drinking, mood, excitability, and stool frequency. Excessive crying	 Parent-completed standardized questionnaire using a 4-point scale (less than 1h a week; 2–6h a week; 1–4h a day; more than 5h a day). Standardised questionnaire using 3-point scale (more/better, same, less/worse, or don't know. Standardized questionnaire according to Wurmser *et al.,* (2001).
Measles	Purse, 1961 No information available	 No information available

Plagiocephaly	Lessard *et al.,* 2011 Anthropometric measurements Plagiocephalometric measurements Mobility, vitality and positioning of different anatomical structures	 Digital caliper Cranial circumference moulds using a low temperature thermoplastic material (sansplint Osteopathic evaluations
Chronic constipation	Quist and Duray, 2007 – case study (n=1) Leg length measurement Range of Motion Spinal examination Bowel function change	 Physical evaluation Palpation Palpation Patient and parental report
	Tarsuslu *et al.,* 2009 Constipation in patients with cerebral palsy (n=13) Motor development Constipation severity Functional independence Muscle tone	 Gross Motor Functional Classification System Constipation Assessment Scale Functional independence Measure for Children Modified Ashworth Scale
Anthropometric dimensions in adolescent males with low back pain	Ebrall 1994 Sitting height Standing height Pelvic height Suprapelvic height	 Not disclosed Not disclosed Not disclosed Not disclosed
Safety studies N=14	Rageot, 1968; Dupeyron *et al.,* 2003: Survey; Hayes and Bezilla, 2006: Retrospective case note review; Humphreys *et al.,* 2010: Review;	

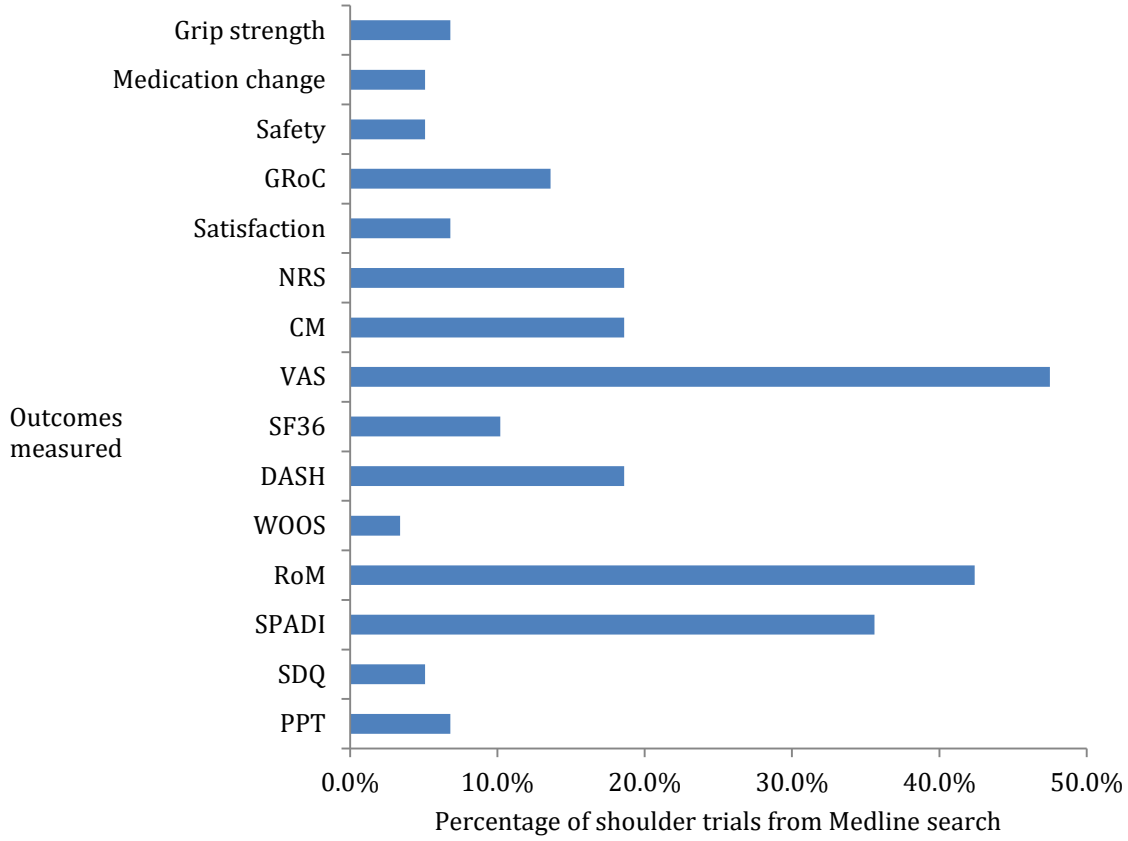
	Jacobi *et al*., 2001: Case report; Klougart *et al.,* 1996: Survey; Miller and Benfield, 2008: Retrospective case note review; Senstad *et al.,* 1996: Survey of clinicians; Senstad *et al.,* 1997: Prospective survey; Shafir and Kaufman, 1992: case study; Sperry and Pfalzgraf, 1990: case study; Todd *et al.,* 2015: systematic review; Vohra *et al.,* 2007: systematic review; Vohra *et al*., 2009: surveillance survey.	

Appendix 7.2 Most frequently cited PROMs in clinical trials involving manual therapy management of hip disorders



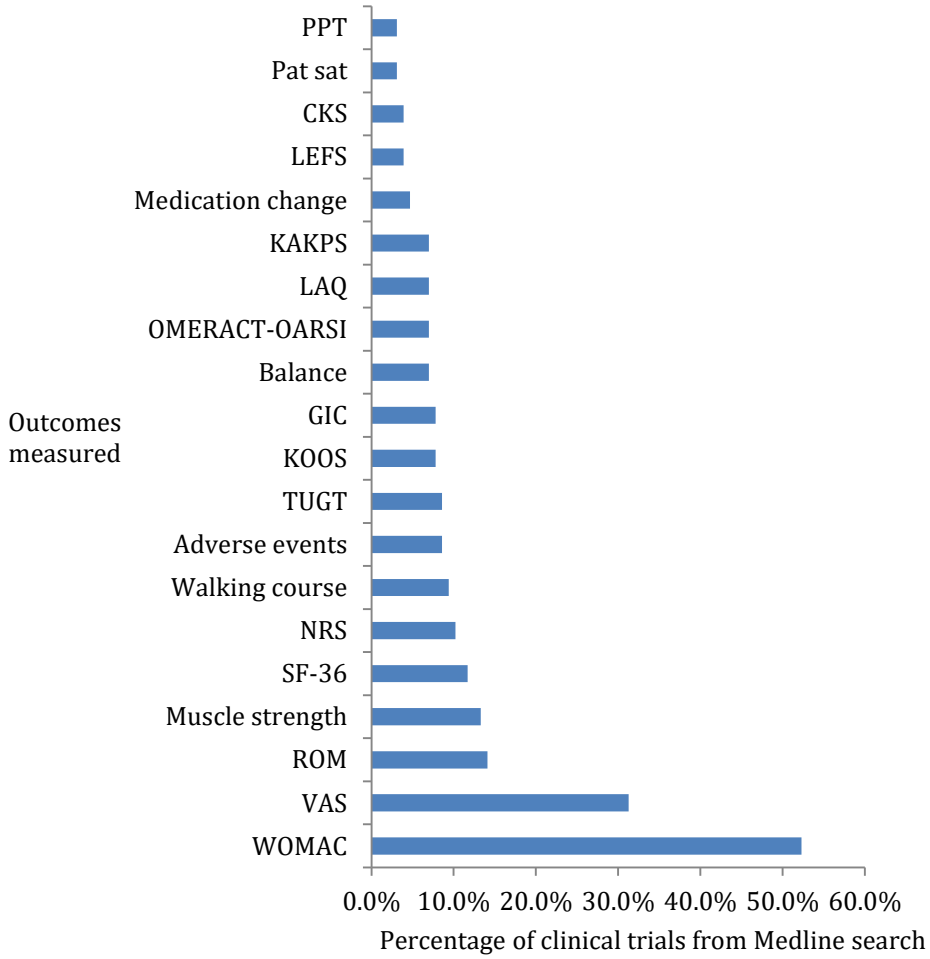
HHS	Harris Hip Score	VAS	Visual Analogue Scale
PASE	Physical Activity Scale for the Elderly	NRS	Numerical Rating Scale
SF36	Short Form 36	RoM	Range of Motion
EQ5D	EuroQoL 5D	TUG	Timed Up and Go
GPE	Global Perceived Effect	BBS	Berg Balance Scale
HOOS	Hip Osteoarthritis Outcome Score	WOMAC	Western Ontario and McMasters Universities Osteoarthritis Index

Appendix 7.3 Most frequently cited PROMs in clinical trials involving manual therapy management of shoulder disorders



RoC	Global Rating of Change	RoM	Range of Motion
NRS	Numerical Rating Scale	SPADI	Shoulder Pain and Disability Index
VAS	Visual Analogue Scale	SDQ	Shoulder Disability Questionnaire
SF36	Short Form 36	PPT	Pain Pressure Threshold
DASH	Disabilities of the Arm, Shoulder, and Hand	CM	Constant-Murley
WOOS	Western Ontario Arthritis of the Shoulder Index	ASES	American Shoulder and Elbow Society Scoring System

Appendix 7.4 Most frequently cited PROMs in clinical trials involving manual therapy management of knee disorders



PPT	Pain Pressure Threshold	KOOS	Knee injury and Osteoarthritis Outcome Score
CKS	Cincinnati Knee Scale	TUGT	Timed Up and Go Test
LEFS	Lower Extremity Functional Scale	NRS	Numerical Rating Scale
KAKPS	Kujala Anterior Knee Pain Scale	SF36	Short Form 36
LAQ	Long Arc Quadriceps	ROM	Range of Motion
GIC	Global Impression of Change	VAS	Visual Analogue Scale
WOMAC	Western Ontario and McMasters Universities Osteoarthritis Index		

References

Aaronson, N.K., Ahmedzai, S., Bergman, B. (1993) The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute* 85:365–376.

Accorsi, A., Lucci, C., Di Mattia, L., Granchelli, C., Barlafante, G., Fini, F., Pizzolorusso, G., Cerritelli, F., Pincherle, M. (2014). Effect of osteopathic manipulative therapy in the attentive performance of children with attention-deficit/hyperactivity disorder. *J Am Osteopath Assoc*. 114(5): 374-381

Aday, L. (1996) *Designing and conducting health surveys*. Jossey-Bass: San Francisco.
Advertising Standards Authority (ASA) (2016a) Retrieved 16th June, 2016 from https://www.asa.org.uk/About-ASA.aspx

Advertising Standards Authority (ASA) (2016b) Relevant Code Rule (Rule 21.1). Retrieved 16th June, 2016 from https://www.asa.org.uk/Rulings/Adjudications/Display-Code.aspx?CodeId=%7BE8E95401-CED5-4104-900D-FD5CF82BDCFC%7D&ItemId=%7B03AFF62C-C54E-4351-B824-A0388A9F8C5B%7D

Airaksinen, O., Brox, J.I., Cedraschi, C., Hildebrandt, J. (2006) European guidelines for the management of chronic nonspecific low back pain. *European Spine Journal* 15 (Suppl 2):S192-S300.

Aitken, R.C. (1969) Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society of Medicine* 62:989-993.

Ajzen, I. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Ajzen, I. (1988). *Attitudes, personality, and behavior*. Milton Keynes, CA: Open University Press.

Alcantara, J., Davis, J. (2010). The chiropractic care of children with attention-deficit/hyperactivity disorder: a retrospective case series. *Explore* (NY). 6(3): 173-182.

Allwood D, Hildon Z, Black N. (2013) Clinicians' views of formats of performance comparisons. *Journal of Evaluation in Clinical Practice* 19:86-93.

Altman, D.G. (1991) *Practical Statistics for Medical Research.* Chapman & Hall: London.

American Marketing Association (AMA). Definition of a marketing programme. Retrieved May 20th, 2016 from www.ama.org.

Andersen, R.S., Hansen, R.P., Søndergaard, J., Bro, F. (2008) Learning based on patient case reviews: an interview study. *BMC Medical Education* 8:43.

Andersson, G.B., Lucente, T., Davis, A.M., Kappler, R.E., Lipton, J.A., Leurgans, S. (1999). A comparison of osteopathic spinal manipulation with standard care for patients with low back pain. *New England Journal of Medicine* 341(19):1426-31.

Andresen, E.M., Meyers, A.R. (2000) Health-related quality of life outcomes measures. *Archives of Physical Medicine and Rehabilitation* 81:S30–S45.

Ang, D.C., Bair, M.J., Damush, T.M., Wu, J., Tu, W., Kroenke, K. (2010). Predictors of pain outcomes in patients with chronic musculoskeletal pain co-morbid with depression: results from a randomised controlled trial. *Pain Medicine*, 11(4):482-491.

Anonymous (1962) Towards a gold standard. *Lancet* 279(7226):418.

Anonymous. (1989) Working for patients. NHS White Paper. New diagnosis--new prescription. *The Health Service Journal* 99(5136):134-7.

Anonymous. The Patient Reported Outcomes Measurement Information System (PROMIS): A walk through the first four years. (2009,1-16). Retrieved 1st May, 2014 from http://www.nihpromis.org/Documents/PROMIS_The_First_Four_Years.pdf

Any Qualified Provider (AQP). Retrieved 13th May, 2014 from www.supply2tohealth.org.

Any Qualified Provider. Retrieved 2nd June, 2015 from https://online.contractsfinder.businesslink.gov.uk/.

Apple. (2005) https://developer.apple.com/library/content/navigation/index.html#topic=Technical+Q%26amp;As&section=Resource+Types

Appleby, J. How does NHS spending compare with health spending internationally? January, 2016. Retrieved 2nd May, 2016 from http://www.kingsfund.org.uk/blog/2016/01/how-does-nhs-spending-compare-health-spending-internationally.

Arden, N.K., Kiran, A., Judge, A. (2011) What is a good patient reported outcome after total hip replacement? *Osteoarthritis and Cartilage* 19:155-62.

Aristotle. Retrieved 1st May,2016 from http://www.goodreads.com/author/show/2192.Aristotle

Armson, H., Kinzie, S., Hawes, D., Roder, S., Wakefield, J., Elmslie, T. (2007) Translating learning into practice: lessons from the practice-based small group learning program. *Canadian Family Physician* 53(9):1477-85.

Arnold, R.M., Weissman, D.E. (2004) Broaching the topic of a palliative care consultation with patients and families. *Journal of Palliative Medicine* 7(3):472-3.

Arnstein, S. (1969) A ladder of participation in the USA. *Journal of the American Institute of Planners* 35(4):216-224.

Arthritis Care UK (2015). Patient Reported Outcome Measures. Retrieved 2nd June, 2015 from http://www.arthritisresearchuk.org/policy-and-public-affairs/policy-priorities-and-projects/musculosketal-health-services/patient-reported-outcome-measures.aspx.

Arthritis Research Campaign. (2002) Arthritis the big picture. Arthritis Research Campaign.

Artus, M., van der Windt, D.A., Jordan, K.P., Hay, E.M. (2010). Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. *Rheumatology* (Oxford) 49(12):2346-56.

Asch, S.M., Kerr, E.A., Keesey, J., Adams, J.L., Setodji, C.M., Malik, S., McGlynn, E.A. (2006) Who is at greatest risk for receiving poor-quality health care? *N Engl J Med.* 354(11):1147-56.

Assendelft, W.J., Morton, S.C., Yu, E.I. (2003) Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. *Annals of Internal Medicine* 138(11):871-81.

Atkinson, P. (1988) Ethnomethodology: A Critical Review. *Annual Review of Sociology* 14: 441–65.

Atlas, S.J., Deyo, R.A., van den Ancker, M., Singer, D.E., Keller, R.B., Patrick, D.L. (2003) The Maine-Seattle back questionnaire: a 12-item disability questionnaire for evaluating patients with lumbar sciatica or stenosis: results of a derivation and validation cohort analysis. *Spine* 28:1869-1876.

Avis, M., Elkan, R., Patel, S., Walker, B.A., Ankti, N., Bell, C. (2008) Ethnicity and participation in cancer self-help groups. *Psychooncology*. 17(9):940-7.

Badia, X., Baro, E. (2001). Cuestionarios de salud en España y su uso en atención primaria. *Aten Primaria* 28:349–356.

Baker, R. (1998) Measuring change in health in general practice: a comparison of a simple transition question with the Nottingham Health Profile. *International Journal of Quality in Health Care* 10(3):207-12.

Balon, J., Aker, P.D., Crowther, E.R., Danielson, C., Cox, P.G., O'Shaughnessy, D., Walker, C., Goldsmith, C.H., Duku, E., Sears, M.R. (1998). A comparison of active and simulated chiropractic manipulation as adjunctive treatment for childhood asthma. *N Engl J Med.* 339(15): 1013-1020

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall.

Barber, B.L., Santanello, N.C., Epstein, R.S. (1996) Impact of the global on patient perceivable change in an asthma specific quality of life questionnaire. *Quality of Life Research* 5:117-122.

Barna, A., Porat, I.P (1976). *Introduction to Microcomputers and the Microprocessors.* New York: Wiley.

Barnett, K., Mercer, S.W., Norbury, M., Wyke, S., Guthrie, B. (2011) Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet* 380(9836):37-43.

Barry,M., Edgman-Levitan,S. (2012) Shared decision making—the pinnacle of patient-centered care. *New England Journal of Medicine* 366:780-1.

Bastecki, A. V., Harrisson, D.E., Haas, J.W. (2004). Cervical kyphosis is a possible link to attention-deficit/hyperactivity disorder. *J Manipulative Physiol Ther*. 27(8): e14.

Bate, P., Robert, G. (2006) Experience-based design: from redesigning the system around the patient to co-designing services with the patient. *Quality and Safety in Health Care* 15:307 –10.

Bath, B., Grono, S.L. (2015). Biopsychosocial predictors of short term success among people with low back pain referred to a physiotherapy spinal triage service. *Journal of Pain Research*, 8:189-202.

BDUK Market Test Pilots (1st February, 2015). Retrieved 2nd June, 2015 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/409858/15-02-09_BDUK_Summary_of_feasibility_phase_final__1_.pdf.

Beard, D.J., Harris, K., Dawson, J., Doll, H., Murray, D.W., Carr, A.J., Price, A.J. (2015) Meaningful changes for the Oxford hip and kbee scores after joint replacement surgery. *Journal of Clinical Epidemiology* 68:73-79.

Beaton, D.E. (2000) Understanding the relevance of measured change through studies of responsiveness. *Spine* (Phila Pa 1976) 25(24):3192-9.

Beaton, D.E., Bombardier, C., Guillemin, F., Ferraz, M.B. (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186 -91.

Beaton, D.E., Bombardier, C., Katz, J.N., Wright, J.G. (2001) A taxonomy for responsiveness. *Journal of Clinical Epidemiology* 54: 1204-1217.

Beddow, A. Physiotherapy Workforce Review. Centre for Workforce Intelligence, 2010. Retrieved 6th March, 2016 from http://www.csp.org.uk/system/files/secure/cfwi_physiotherapy_workforce_review_2010_0.pdf.

Bell, R., Kravitz, R. (2002) Unmet Expectations for Care and the Patient - Physician Relationship. *Journal of General and International Medicine.* 17: 817-824.

Bellg A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., Czajkowski, S. (2004) Enhancing treatment fidelity in health behaviour change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol*. 23: 443-451.

Beneciuk, J.M., Fritz, J.M., George, S.Z. (2014). The STarT Back Screeing Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings. *Journal of Orthopaedics and Sports Physical Therapy*, 44(9):656-64.

Bennett, D., Walsh, M., O'Sullivan, R., Gallagher, J., O'Brien, T., Newman, C.J. (2007). Use of a dynamic foot pressure index to monitor the effects of treatment for equinus gait in children with cerebral palsy. *J Pediatr Orthop.* 27(3): 288-294.

Benson, H.R. (1994). An introduction to benchmarking in healthcare. *Radiology Management*, 16(4):35-9.

Benson, K., Hartz, A.J. (2000) A comparison of observational studies and randomised controlled trials. *New England Journal of Medicine* 342: 1878–1886.

Beresford, P. (2002) User involvement in research and evaluation: liberation or regulation? *Social Policy and Society* 1(2):95-105.

Beresford, P., Croft, S. (1992) The politics of participation. *Critical Social Policy* 35:20-44.

Bergh, I., Sjöström, B., Odén, A., Steen, B. (2000) An application of pain rating scales in geriatric patients. *Aging* (Milano). 12(5):380-7.

Bergh, I., Sjöström, B., Odén, A., Steen, B. (2001) Assessing pain and pain relief in geriatric patients with non-pathological fractures with different rating scales. *Aging* (Milano). 13(5):355-61.

Bergner, M., Bobbitt, R.A., Carter, W.B., Gibson, B.S. (1981) The Sickness Impact Profile: development and final version of a health status measure. *Medical Care* 19:787-805.

Bergstrom, G., Jensen, I.B., Bodin, L., Linton, S.J., Nygren, A.L. (1998) Reliability and factor structure of the multidimensional pain inventory – Swedish language version (MPI-S). *Pain* 75:101-110.

Bernard, H. (2002) *Research methods in anthropology: Qualitative and quantitative approaches.* Lanham: Altamira.

Bero, L.A., Grilli, R., Grimshaw, J.M., Harvey, E., Oxman, A.D., Thomson, M.A. (1998) Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ.* 317(7156):465-8.

Berwick, D.M., Hackbarth, A.D. (2012) Eliminating waste in US health care. *Journal of the American Medical Association* 307(14):1513-6.

Beyer, J.E., Knott, C.B. (1998) Construct validity estimation for the African-American and Hispanic versions of the Oucher Scale. *J Pediatr Nurs*. 13(1):20-31.

Beurskens, A., de Vet, H., Koke, A. (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 65:71-6.

Bieri, D., Reeve, R.A., Champion, G.D., Addicoat, L., Ziegelr, J.B. (1990) The Faces Pain Scale for the self-assessment of the severity of pain experienced by children:

development, initial validation, and preliminary investigation for ratio scale properties. *Pain* 41:139-50.

Bijur, P.E., Silver, W., Gallagher, E.J. (2001) Reliability of the visual analogue scale for measurement of acute pain. *Academic Emergency Medicine* 8:1153-1157.

Bilyayeva, T.A. (2012) Cross-Cultural Comparative Study of Users' Perception of the Navigation Organization of an E-Commerce Web Application. Masters Thesis USA; University of North Florida.

Binns, H.J.,Lanier, D., Pace, W.D. (2007) Describing primary care encounters, the Primary Care Network Survey and the National Ambulatory Medical Care Survey. *Annals of Family Medicine* 5(1):39-47.

Bird, M-L., Callisaya, M.L., Cannell, J., Gibbons, T., Smith, S.T., Ahuja, K.D.K. (2016). Accuracy, Validity, and Reliability of an Electronic Visual Analog Scale for Pain on a Touch Screen Tablet in Healthy Older Adults: A Clinical Trial. *Interactive Journal of Medical Research* 5(1):e3 i-viii.

Bishop, F., Yardley, L., Lewith, G. (2007) A systematic review of beliefs involved in the use of complementary and alternative medicine. *Journal of Health Psychology*. 12(6): 851-67

Bishop, F., Yardley, L., Lewis, G. (2010) A within-subjects trial to test the equivalence of online and paper-based outcome measures: the Roland Morris Disability Questionnaire. *BMC Musculoskeletal Disorders* 11:113.

Bishop, F.L., Barlow, F., Coghlan, B., Lee, P., Lewith, G.T. (2011a) Patients as healthcare consumers in the public and private sectors: a qualitative study of acupuncture in the UK. *BMC Health Serv Res*. 27;11:129

Bishop, F.L., Massey, Y., Yardley, L., Lewith, G.T. (2011b) How patients choose acupuncturists: a mixed-methods project. *J Altern Complement Med*. 17(1):19-25.
Black, N., Jenkinson, C. (2009) Measuring patients' experiences and outcomes. *British Medical Journal* 339:2495-2504.

Black, N., Varagunam, M., Hutchings, A. (2014) Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Quality and Safety* 23(7):534-42.

Blomerth, P. R. (1994). Functional nocturnal enuresis. *J Manipulative Physiol Ther*. 17(5): 335-338.

Blount, K.J., Krompringer, W.J., Maljianian, R. (2012) Moving towards a standard for spinal fusion outcomes assessment. *Journal of Spinal Disorders and Technology* 15:16-23.

Blum-Fowler, C., Peterson, C., McChurch, J.F., Le Clech, Y., Humphreys, B.K. (2013) Translation and validation of the German version of the Bournemouth questionnaire for low back pain. *Chiropr Man Therap*. (1):32.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method* (1st ed.). Berkeley: University of California Press.

Bobo, L., Wiomedu, R., Knox, A. (1991) Principles of intercultural medicine in an internal medicine program. *American Journal of Medical Science* 302(4):244-8.

Boden, S.D., McCowin, P.R., Davis, D.O., Dina, T.S., Mark, A.S., Wiesel, S. (1990) Abnormal magnetic resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation. *Journal of Bone and Joint Surgery* 72:1178-1184.

Bodenheimer, T. (1999) The American health care system--physicians and the changing medical marketplace. *N Engl J Med*. 340(7):584-8.

Boers, M., Idzerda, L., Kirwan, .JR., Beaton, D., Escorpizo, R., Boonen, A., Magasi, S., Sinha, I., Stucki, G., Tugwell, P. (2014) Toward a generalized framework of core measurement areas in clinical trials: a position paper for OMERACT 11. *J Rheumatol.* 41(5):978-85.

Boissy, P., Jacobs, K., Roy, S.H. (2006) Usability of a barcode scanning system as a means of data entry on a PDA for self-report health outcome questionnaires: a pilot study in individuals over 60 years of age. *BMC Medical Informatics and Decision Making* 6:42.

Bolin, D. J. (2010). The application of osteopathic treatments to pediatric sports injuries. *Pediatr Clin North Am*. 57(3): 775-794

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* 8:503-10.

Bombardier, C., Tugwell, P. (1987) Methodological considerations in functional assessment. *Journal of Rheumatology*. Supplement. 14 Suppl 15:6–10.

Bombardier, C., Kerr, M.S., Shannon, H.S., Frank, J.W. (1994) A guide to interpreting epidemiologic studies on the etiology of back pain. *Spine* (Phila Pa 1976) 19(18 Suppl):2047S-2056S.

Bombardier, C. (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* (Phila Pa 1976) 25(24):3100-3.

Bornmann, L., Mutz, R. (2014) Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references: Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References. *Journal of the Association for Information Science and Technology* 66(11): 2215-2222.

Bourdieu, P. (2004). *Science of science and reflexivity*. Cambridge, UK: Polity Press.

Bowen, G.A. (2009) Supporting a Grounded Theory with an Audit Trail: An Illustration. *International Journal of Social Research Methodology*, 12(4):305 - 316.

Boyce, M.B., Browne, J.B., Greenhalgh, J. (2014) The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research *BMJ Quality and Safety* 23:6 508-518.

Brand, P. L., Engelbert, R.H., Helders, P.J., Offringa, M. (2005). Systematic review of the effects of therapy in infants with the KISS-syndrome (kinetic imbalance due to suboccipital strain). *Ned Tijdschr Geneeskd*. 149(13): 703-707.

Bream, E., Charman, S.C., Clift, B. (2010). Relationship between patients' and clinicians' assessments of health status before and after knee arthroplasty. *Qual Saf Health Care* 9:1-3.

Breckenridge, K., Bekker, H.L., Gibbons, E., van der Veer, S.N., Abbott, D., Briançon, S., Cullen, R., Garneata, L., Jager, K.J., Lønning, K., Metcalfe, W., Morton, R.L., Murtagh, F.E., Prutz, K., Robertson, S., Rychlik, I., Schon, S., Sharp, L., Speyer, E., Tentori, F., Caskey, F.J. (2015) How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. *Nephrology, Dialysis, and Transplantation* 30(10):1605-14.

Breen, A., Breen, R. (2003). Back pain and satisfaction with chiropractic treatment: what role does physical outcome play? *Clinical Journal of Pain*, 19:263-268.

Breivik, H., Borchgrevink, P.C., Allen, S.M., Rosseland, L.A. (2008) Assessment of Pain. *British Journal of Anaesthesia* 101(1):17-24.

Brehaut, J.C., Poses, R., Shojania, K.G., Lott, A., Man-Son-Hing, M., Bassin, E., Grimshaw, J. (2007) Do physician outcome judgments and judgment biases contribute to inappropriate use of treatments? Study protocol. *Implement Sci*. 2007 Jun 7;2:18.

Briggs, M., Closs, J.S. A descriptive study of the use of visual analogue scale and verbal rating scales for the assessment of postoperative pain in orthopaedic patients. *Journal of Pain and Symptom Management*, 18:438-446, 1999.

British Osteopathic Association (BOA). (2009) *Stripped Back: An investigation into the 'postcode lottery' of the availability of osteopathy on the NHS in England*. BOA.

Bronfort, G., Evans, R.L., Kubic, P., Filkin, P. (2001). Chronic pediatric asthma and chiropractic spinal manipulation: a prospective clinical series and randomized clinical pilot study. *J Manipulative Physiol Ther*. 24(6): 369-377.

Brønfort, G., Nilsson, N., Haas, M. (2004) Non-invasive physical treatments for chronic/recurrent headache. *Cochrane Database of Systematic Reviews* Issue 3. Art. No.: CD001878. DOI:10.1002/14651858.CD001878.pub2.

Brønfort, G., Haas, M., Evans, R., Leininger, B., Triano, J. (2010) Effectiveness of manual therapies: the UK evidence report. *Chiropractic and Osteopathy* 18:3.

Bronner, S., Agraharasamakulam, S., Ojofeitimi, S. (2010) Reliability and validity of a new ankle electrogoniometer. *Journal of Medical Engineering and Technology* 2010;34(5-6):350-5.

Bucknall, T., Fossum, M. (2015) It Is Not That Simple nor Compelling! Comment on "Translating Evidence Into Healthcare Policy and Practice: Single Versus Multi-faceted Implementation Strategies - Is There a Simple Answer to a Complex Question?". *Int J Health Policy Manag.* 4(11):787-8.

Bull, L. (2007). Sunflower therapy for children with specific learning difficulties (dyslexia): a randomised, controlled trial. *Complement Ther Clin Pract*. 13(1): 15-24.

Burton, A.K. (1981) Back pain in osteopathic practice. *Rheumatology and Rehabilitation* 20: 239-46.

Burton, A.K., McClune, T.D., Clarke, R.D., Main, C.J. (2004). Long-term follow-up of patients with low back pain attending for manipulative care: outcomes and predictors. *Manual Therapy*, 9:30-35.

Burton, A.K., Tillotson, K.M., Cleary, J. (2009) Single-blind randomized controlled trial of chemonucleolysis and manipulation in the treatment of symptomatic lumbar disc herniation. *European Spine Journal* 9:202-7.

Burton, L. Low income and digital exclusion, 2013. Retrieved 23rd March, 2014 from http://www.poverty.ac.uk/editorial/low-income-and-digital-exclusion.

Bushnell, D.M., Martin, M.L., Parasuraman, B. (2003) Electronic versus paper questionnaires: a further comparison in persons with asthma. *Journal of Asthma* 40(7):751-62.

Caldicott, F. The Information Governance Review. Information: To share or not to share? March, 2013. Retrieved 13th May, 2014 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf.

Caldwell, D.F., Chatman, C., O'Reilly, C.A., Ormiston, M., Lapiz, M. (2008) Implementing strategic change in a health care system: The importance of leadership and change readiness. *Health Care Management Review* 33(2): 124-133.

Cameron, E. C., Maehle, V., Reid, J. (2005). The effects of an early physical therapy intervention for very preterm, very low birth weight infants: a randomized controlled clinical trial. *Pediatr Phys Ther*. 17(2): 107-119.

Cameron, J.J., McColl, M.A. (2015) Learning client-centred practice short report: experience of OT students interacting with "expert patients. *Scandinavian Journal of Occupational Therapy* 22(4):322-4.

Campbell, R.J. (2007) Change Management in Health Care. *The Health Care Manager* 27(1): 23–39.

Canadian Task Force on the Periodic Health Examination. (1979) The periodic health examination. *Canadian Medical Association Journal* 121, 1193–1254.

Care Act (2014). Retrieved 13th June, 2014 from http://services.parliament.uk/bills/2013-14/care.html.

Care Response programme. Retrieved 1st May, 2014 from www.Care_Response.org.uk.

Carnes, D. (2006) *Understanding and measuring chronic musculoskeletal pain in the community using self-completed pain drawings*. PhD thesis, 240.

Carnes, D., Mullinger, B., Underwood, M. (2010) Defining adverse events in manual therapies: A modified Delphi consensus study. *Manual Therapy* 15(1):2-6.

Carnes, D., Mars, T.S., Mullinger, B., Froud, R., Underwood, M. (2010) Adverse events and manual therapy: A systematic review. *Manual Therapy* 15(4):355-63.

Carnes, D., Fawkes, C. (2012) What are the differences between osteopathy, physiotherapy and chiropractic? *The Osteopath* 15(4):17-19.

Carpenter, J.S., Brockopp, D. (1995) Comparison of patients' ratings and examination of nurses' responses to pain intensity rating scales. *Cancer Nurs*. 18(4):292-8.

Carr, E.C., Worth, A. (2001) The use of the telephone interview for research. *NT Research* 6:511–524.

Carr-Hill, R. (1992) The measurement of patient satisfaction. *British Medical Journal*. 14(236-49)

Carreon, L.Y., Glassman, S.D., Howard, J. (2008). Fusion and nonsurgical treatment for symptomatic lumbar degenerative disease: a systematic review of Oswestry Disability Index and MOS Short Form-36 outcomes. *Spine* Journal 8(5):747-55.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., Balain, S. (2007) A conceptual framework for implementation fidelity. *Implementation Science* 2:40

Catlett, E. (1937). Retrieved 30th May, 2016 from http://www.nmwa.org/works/singing-their-songs.

Cerritelli, F., Pizzolorusso, G., Renzetti, C., D'Incecco, C., Fusilli, P., Perri, P.F., Tubaldi, L., Barlafante, G. (2013). Effectiveness of osteopathic manipulative treatment in neonatal intensive care units: protocol for a multicentre randomised clinical trial. *BMJ Open*. 3(2).(pii): e002187

Chalmers, T.C., Celano, P., Sacks, H.S., Smith, H. (1983) Bias in treatment assignment in controlled clinical trials. *New England Journal of Medicine* 309: 1358–1361.

Chalmers, I., Hedges, L.V., Cooper, H. (2002) A brief history of research synthesis. *Evaluation and the Health Professions*. 25(1):12-37

Chambers English Dictionary (1989), Cambridge, UK: Chambers.

Chambers English Dictionary (2014), Cambridge, UK: Chambers.

Chapman, J.R., Norvell, D.C., Hermsmeyer, J.T., Bransford, R.J., DeVine, J., McGirt ,M.J., Lee, M,J. (2011). Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine* (Phila Pa 1976). 36(21 Suppl):S54-68.

Chapple, A. (1999) The use of telephone interviewing for qualitative research. *Nurse Researcher* 6: 85-93.

Chartered Society of Physiotherapists (2012). *Developing a Musculoskeletal Patient Reported Outcome Tool: Report from a Community Discussion Workshop.*

Chassany, O., Sagnier, P., Marquis, P., Fullerton, S., Aaronson, N.K., for the European Regulatory Issues on Quality of Life Assessment Group (2002). Patient-reported outcomes: the example of health-related quality of life - a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. *Drug Information Journal* 36:209–238.

Chen, J., Ou, L., Hollis, S.J. (2013). A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Services Research*, 13:211

Cheng, J. C., Wong, M.W., Tang, S.P., Chen, T.M., Shum, S.L., Wong, E.M. (2001). Clinical determinants of the outcome of manual stretching in the treatment of congenital muscular torticollis in infants. A prospective study of eight hundred and twenty-one cases. *J Bone Joint Surg Am*. 83-A(5): 679-687.

Cheshire, A., Polley, M., Peters, D., Ridge, D. (2013) Patient outcomes and experiences of an acupuncture and self-care service for persistent low back pain in the NHS: a mixed methods approach. *BMC Complement Altern Med*. 13:300.

Chew-Graham, C., Hunter, C., Langer, S. (2013) How QOF is shaping primary care review consultations: a longitudinal qualitative study. *BMC Family Practice* 14:103.

Chiarotto, A., Deyo, R.A., Terwee, C.B., Boers, M., Buchbinder, R., Corbin, T.P., Costa, L.O., Foster, N.E., Grotle, M., Koes, B.W., Kovacs, F.M., Lin, C.W., Maher, C.G., Pearson, A.M., Peul, W.C., Schoene, M.L., Turk, D.C., van Tulder, M.W., Ostelo, R.W. (2015) Core outcome domains for clinical trials in non-specific low back pain. *European Spine Journal* 24(6).

Chiarotto, A., Maxwell, L.J., Terwee, C.B., Wells, G.A., Tugwell, P., Ostelo, R.W. (2016). Roland-Morris Disability Questionnaire and Oswestry Disability Index: Which Has Better Measurement Properties for Measuring Physical Functioning in Nonspecific Low Back Pain? Systematic Review and Meta-Analysis. *Physical Therapy* Apr 14. [Epub ahead of print].

Chila, A.G., Jeffries, F.F., Levin, S.M. (1990) Is manipulation for your practice? *Patient Care* 77-92.

Chown, M., Whittamore, L., Rush, M. (2008) A prospective study of patients with chronic back pain randomised to group exercise, physiotherapy or osteopathy. *Physiotherapy* 94:21-28.

Christensen, L., Mendoza, J. (1986) A method of assessing change in a single subject: an alteration of the RC index. *Behavioural Therapy* 17:305-8.

Cipriani, C., Segil, J., Birdwell, J., Weir, R. (2014) Dexterous control of a prosthetic hand using fine-wire intramuscular electrodes in targeted extrinsic muscles". *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 1–1.

Clancy, C.M., Eisenberg, J.M. (1998) Outcomes research: measuring the end results of health care. *Science* 282:245-246.

Clar, C., Tsertsvadze, A., Court, R. (2014) Clinical effectiveness of manual therapy for the management of musculoskeletal and non-musculoskeletal conditions: systematic review and update of UK evidence report. *Chiropractic and Manual Therapies* 22(1):12.

Clare, A. (1980) Psychiatry in Dissent. London; Routledge.

Claassen, J.A.H.R. (2005) The gold standard: not a golden standard. *British Medical Journal* 330:1121.

Cleeland, C.S., Ryan, M.K. (1994) Pain assessment: global use of the Brief Pain Inventory. *Annals of Academic Medicine* 23:129-38.

Clement, R.C., Welander, A., Stowell, C., Cha, T.D., Chen, J.L., Davies, M., Fairbank, J.C., Foley, K.T., Gehrchen, M., Hagg, O., Jacobs, W.C., Kahler, R., Khan, S.N., Lieberman, I.H., Morisson, B., Ohnmeiss, D.D., Peul, W.C., Shonnard, N.H., Smuck, M.W., Solberg, T.K., Stromqvist, B.H., Van Hooff, M.L., Wasan,A.D., Willems, P.C., Yeo, W., Fritzell, P. (2015) A proposed set of metrics for standardized outcome reporting in the management of low back pain**.** *Acta Orthop*. 86(5): 523–533.

Clinical Standards Advisory Group (CSAG) (1994). *Back pain*. London: HMSO

Cochrane, A.L. (1972). *Effectiveness and Efficiency: Random Reflections on Health Services*. Oxford: Nuffield Provincial Hospitals Trust.

Coelho, R.A., Siqueira, F.B., Ferreira, P.H. (2008) Responsiveness of the Brazilian-Portuguese version of the Oswestry Disability Index in subjects with low back pain. *European Spine Journal* 17(8):1101-1106

Cohen, J. (1977) *Statistical power analysis for the behavioural sciences*. New York: Academic Press.

Cohen, L., Manion, L. (2000). *Research methods in education* (5th edition). Oxford, UK: Routledge.

Cohen, L., Manion, L., Morrison, K. (2011) *Research Methods in Education* (7th edition): Oxford, UK: Routledge.

Cohn, F.J. (1856) Beobachtungen über den Ban and die Fortpflanzung von *Volvox globator. Ubersicht der Arbeiten and Veranderungen der Schlesischen Gesllschaft fur vaterlandsiche Kultur* 77–83.

Cohn, S.P. (2006) *Privacy and confidentiality in the nationwide health information network*. Washington: National Committee on Vital and Health Statistics.

Colin-Thom é, D. (2009). *Mid Staffordshire NHS Foundation Trust: A review of lessons learnt for commissioners and performance managers following the Healthcare Commission investigation*. London: Department of Health.

Collins-Fulea, C., Mohr, J.J., Tillett, J. (2005). Improving midwifery practice: the American College of Nurse-midwifery benchmarking project. *Journal of Midwifery and Women's Health*, 50(6):461-471.

Concato, J., Shah, N.,Horwitz, R.I. (2000) Randomised controlled trials, observational studies and the hierarchy of research designs. *New England Journal of Medicine* 342: 1887–1892.

Cook, C.E., Arnold, P.M., Passias, P.G., Frempong-Boadu, A.K., Isaacs, R. (2015). Predictors of pain and disability outcomes in one thousand, one hundred and eight patients who underwent lumbar discectomy surgery. *International Orthopaedics*, 39)11):2143-51.

Cook, C., Heath, F., Thompson, R.L. (2000) A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement* 60(6):821-836.

Cook, K.F., Choi, S.W., Crane, P.K., Deyo, R.A., Johnson, K.L., Amtmann, D. (2008) Letting the cat out of the bag: Comparing computer adaptive tests and an 11-item short form version of the Roland-Morris Disability Questionnaire. *Spine* 33(12): 1378-1383.

Cook, K. S., Emerson, R. M. (1978). Power, equity and commitment in exchange networks. *American Sociological Review*, 43: 721-739

Cooke, T., Watt, D., Wertzler, W. (2006) Patient expectations of emergency department care: Phase II - a cross-sectional survey. *Canadian Journal of Emergency Medicine*. 8(3):148-157.

Coomber, R. (1997). Using the Internet for survey research. *Sociological Research Online*, 2(2).

Cope, D.G. (2014) Computer-assisted qualitative data analysis software. *Oncology Nursing Forum* 41(3):322-3.

Copeland, J.M., Taylor, W.J., Dean, S.G. (2008) Factors influencing the use of outcome measures for patients with low back pain: a survey of New Zealand physical therapists. *Physical Therapy* 88(12):1492–1505.

Core Outcome Measures in Effectiveness Trials (COMET). Retrieved 6th June, 2015 from http://www.comet-initiative.org/.

COSMIN (COnsensus-based Standards for the selection of health Measurement Instruments) checklist. Retrieved 1st August, 2013 from http://www.cosmin.nl/images/upload/File/checklist%20final%20feb%202010.pdf.

COSMIN manual checklist for assessment of studies of measurement properties of instruments. Retrieved 20th May, 2015 from http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf.

COSMIN taxonomy. Retrieved 12th January, 2016 from http://www.cosmin.nl/images/upload/files/COSMIN%20taxonomy.pdf.

Coulter, A. (1999) Paternalism or partnership? *British Medical Journal* 319:719.

Coulter, A. (2005) What do patients and the public want from primary care? *British Medical Journal* 331(7526): 1199-1201.

Coulter, A., Fitzpatrick, R., Cornwell, J. (2009) *The Point of Care Measures of patients' experience in hospital: purpose, methods and uses.* Retrieved 6th June, 2015 from https://www.kingsfund.org.uk/sites/files/kf/Point-of-Care-Measures-of-patients-experience-in-hospital-Kings-Fund-July-2009.pdf.

Coulter, A., Collins, A. (2011) Making shared decision-making a reality: no decision about me, without me. Retrieved 1st May, 2014 from

http://www.kingsfund.org.uk/sites/files/kf/Making-shared-decision-making-a-reality-paper-Angela-Coulter-Alf-Collins-July-2011_0.pdf.

Couper, M.P. (2005) Technology trends in survey data collection. *Social Science Computer Review* 23(4):486-501.

Couper, M.P., Tourangeau, R., Conrad, F.G., Singer, E. (2006) Evaluating the effectiveness of visual analogue scales: a web experiment. *Social Science Computer Review* 24:227.

Cox, K., Bergen, A., Norman. I.J. (1993) Exploring consumer views of care provided by the Macmillan nurse using the critical intent technique. *Journal of Advanced Nursing* 18: 408-15.

Cramer, G.D., Cambron, J., Cantu, J.A., Dexheimer, J.M., Pocius, J.D., Gregerson, D., Fergus, M., McKinnis, R., Grieve, T.J. (2013) Magnetic resonance imaging zygapophyseal joint space changes (gapping) in low back pain patients following spinal manipulation and side-posture positioning: a randomized controlled mechanisms trial with blinding. *J Manipulative Physiol Ther.* 36(4):203-17.

Crawford, M.J., Rutter, D., Manley, C. (2002) Systematic review of involving patients in the planning and development of health care. *British Medical Journal* 325(7375):1263.

Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Creswell, J.W. (2003) *Research design, qualitative, quantitative, and mixed methods approaches* (Second edition). Thousand Oaks, California, USA: Sage

Croft, P., Papageorgiou, A., McNally, R. (1997). *Low Back Pain - Health Care Needs Assessment.* Oxford: Radcliffe Medical Press.

Cropanzano, R., Mitchell, M.S. (2005) Social Exchange Theory: An Interdisciplinary Review. *Journal of Management* 31(6): 6 874-900.

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.

Crow, E., Gage, H., Hampson, S. (2002) The role of expectancies in the placebo effect and their use in the delivery of health care: A systematic review. *Health Technology Assessment.* 3(3):1-96.

Cusumano, M.A., Yoffie, D.B. (1998) *Competing on Internet Time: Lessons from Netscape and Its Battle with Microsoft*, New York: Free Press/Simon & Schuster.

Dale, O., Hagen, K.B. (2007) Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data. *Journal of Clinical Epidemiology* 60(1):8-17.

Dane, A., Schneider, B. (1998) Program integrity in primary and early secondary prevention: Are implementation effects out of control. *Clin Psychol Rev*. 18: 23-45.

Darzi of Denham, L. (2008). *High quality care for all: NHS Next Stage Review final report.* London: Department of Health.

Data Protection Act. (1998) Retrieved 21-02-2014 from http://www.legislation.gov.uk/ukpga/1998/29/contents.

Davidson, M. (2009) Rasch analysis of 24-, 18- and 11-item versions of the Roland-Morris Disability Questionnaire. *Quality of Life Research* 18:473-481

Davidson, M., Keating, J.L. (2002) A comparison of five low back disability questionnaires: Reliability and responsiveness. *Physical Therapy* 82(1):8-24.

Davies, E., Cleary, P.D. (2004) Hearing the patient's voice? Factors affecting the use of patient survey data in quality improvement. *Quality and Safety in Health Care* 14:428–32.

Davies, E., Shaller, D., Edgman-Levitan, S. (2008) Evaluating the use of a modified CAHPS survey to support improvements in patient-centred care: lessons from a quality improvement collaborative. *Health Expectations* 11(2):160 –76.

Davis, D. (2006) Continuing education, guideline implementation, and the emerging transdisciplinary field of knowledge translation. JCEHP 26(1):5-12.

Davis, M. F., Worden, K., Clawson, D., Meaney, F.J., Duncan, B. (2007). Confirmatory factor analysis in osteopathic medicine: fascial and spinal motion restrictions as correlates of muscle spasticity in children with cerebral palsy. *J Am Osteopath Assoc*. 107(6): 226-232.

Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., Carr, A.J. (2010) The routine use of patient reported outcome measures in healthcare settings. *BMJ.* 340:c186.

DCMS (Department for Culture, Media, and Sport). (2014) Retrieved 21st February, 2016 from https://www.gov.uk/broadband-delivery-uk#history.

de Almeida, R.S., Bourliataux-Lajoinie, S., Martins, M. (2015). Satisfaction measurement instruments for healthcare service users: a systematic review. *Cadernos de Saúde Pública* 31(1):11-25.

De Brún, C., Pearce-Smith, N. (2009) *Searching skills toolkit*. Oxford: Wiley-Blackwell.
Declaration of Helsinki. World Medical Association (2013) Retrieved 30th May, 2016 from http://www.wma.net/en/30publications/10policies/b3/.

Degenhardt, B.F., Kuchera, J.L. (1996) Update on osteopathic medicine concepts and the lymphatic system. *Journal of the American Osteopathic Association* 96:97-100.

Degenhardt, B. F., Kuchera, M.L. (2006). Osteopathic evaluation and manipulative treatment in reducing the morbidity of otitis media: a pilot study. *J Am Osteopath Assoc*. 106(6): 327-334.

Dehling, T., Gao, F., Schneider, S., Sunyaev, A. (2015) Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. *Journal of Medical Internet Research MHealth UHealth* 3:e8.

de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research*, 22(4), 464–474.

Demmelmaier, I., Asenlöf, P., Lindberg, P., Denison, E. (2010). Biopsychosocial predictors of pain, disability, health care consumption, and sick leave in first-episode and long-term back pain: a longitudinal study in the general population. (2010). *International Journal of Behavioural Medicine*, 17(2):79-89.

Denzin, N. (2006). *Sociological Methods: A Sourcebook.* London:EDS Publications Limited.

Denzin, N.K. Lincoln, Y. S., eds. (2005). *The Sage Handbook of Qualitative Research* (3rd ed.). Thousand Oaks, CA: Sage.

Department of Health (1997) *Report on the review of patient-identifiable information*. Department of Health: London. Retrieved 21st May, 2014 from http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4068403

Department of Health (2000). *The NHS Plan. A plan for investment. A plan for reform.* London: The Stationary Office.

Department of Health (2001) *Learning from Bristol: The report of the Public Inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–95*. The Stationery Office: London.

Department of Health (2005). *Research Governance Framework for Health and Social Care* (2nd Edition). Department of Health: London. Retrieved 21st May, 2014 from https://www.gov.uk/government/publications/research-governance-framework-for-health-and-social-care-second-edition.

Department of Health (2006) *Musculoskeletal Services Framework*. Department of Health: London.

Department of Health (2013) *Delivering high quality, effective, compassionate care: Developing the right people with the right skills and the right values*. A mandate from the Government to Health Education England: April 2013 to March 2015. Department of Health: London. Retrieved 21st May, 2015 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/203332/29257_2900971_Delivering_Accessible.pdf

De Solla Price, DJ. (1965) Networks of Scientific Papers. *Science* 149 (3683): 510–515.

Dettori, J.R. (2007) In Chapman JR, Hanson BH, Dettori. *Spine Outcome Measures and Instruments*. Stuttgart: Thieme.

Deutscher, D., Hart, D.L., Dickstein, R., Horn, S.D., Gutvirtz, M. (2008) Implementing an electronic outcome and electronic health record process to create a foundation for clinical practice improvement. *Physical Therapy* 88(2):270-285.

de Vet, H.C., Terwee, C.B., Bouter, L.M. (2003) Current challenges in clinimetrics. *Journal of Clinical Epidemiology* 56(12):1137-41.

de Vet, H.C., Terluin, B., Knol, D.L., Roorda, L.D., Mokkink, L.B., Ostelo, R.W., Hendriks, E.J., Bouter, L.M., Terwee, C.B. (2010) Three ways to quantify uncertainty in individually applied "minimally important change" values. *J Clin Epidemiol.* 63(1):37-45.

de Vet, H., Terwee, C., Mokkink, L., Knol, D. (2011) *Measurement in Medicine*. Cambridge: Cambridge University Press.

de Vet, H. C., Terwee, C., Ostelo, R., and Dekker, Y. (2013). Clinimetrics: Assessing measurement properties of health instruments (C202). The Netherlands, EMGO Institute.

Devkaran, S., O'Farrell, P.N. (2014) The impact of hospital accreditation on clinical documentation compliance: a life cycle explanation using interrupted time series analysis. *BMJ Open*. 4(8):e005240.

Devlin, N.J., Appleby, J. (2010) Getting the most out of proms. London; The Kings Fund.

Deyo, R.A., Centor, R.M. (1986) Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis*. 39(11):897-906.

Deyo, R.A., Diehr, P., Patrick, D.L. (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clinical Trials* 12(4 Suppl):142S-158S.

Deyo, R.A., Patrick, D.L. (1995) The significance of treatment effects: the clinical perspective. *Medical Care* 33(4 Suppl):AS286-91.

Deyo, R.A., Battie, M., Beurskens, A.J., Bombardier, C., Croft, P., Koes, B., Malmivaara, A., Roland, M., Von Korff, M., Waddell, G. (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine* (Phila Pa 1976)23(18):2003-13.

DiGiovanna, E.L., Martinke, D.J., Dowling, D.J. (1991) Introduction to osteopathic medicine. In DiGiovanna EL, Sciowitz S eds. *An Osteopathic Approach to Diagnosis and Treatment*. Philadelphia Pa: JB Lippincott.

Di Lampedusa, G.T. (1960). *The Leopard*. London: Collins Harvill.

Dillman, D.A., Sinclair, M.D., Clark, J.R. (1993) Effects of questionnaire length, respondent friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly* 57:289-304.

Dillman, D.A., Lesser, V., Mason, R., Carlson, J., Willits, F., Robertson, R. (2007). Personalisation of mail surveys for general public and populations with a group identity: results from nine studies. *Rural Sociology* 72(4):632-646.

Dionne, C.E., Koepsell, T.D., Von Korff, M. (1997) Predicting long term functional limitations among back pain patients in primary care settings. *Journal of Clinical Epidemiology* 50:31-43.

Dissing, K. B., Hartvigsen, J., Wedderkopp, N., Hestbæk, L. (2016). Conservative care with or without manipulative therapy in the management of back and neck pain in Danish children aged 9-15. Study protocol for a randomized controlled trial. *Chiropr Man Therap.* 24:5.

Dixon, J.S. (1986) Agreement between horizontal and vertical visual analogue scales. *British Journal of Rheumatology* 25:415-416.

Dobson, L., Cook, T. (1980) Avoiding Type III error in program evaluation: results from a field experiment. *Evaluation and Program Planning*. 3: 269-276.

Doll, R., Peto, R. (1981) *The causes of cancer*. New York: Oxford University Press.

Dommeyer, C.J., Baum, P., Hanna, R.W., Chapman, K.S. (2004) Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment and Evaluation in Higher Education* 29(5):611-623.

Dong, H. Y. (2009). Efficacy of manipulative therapy for night waking in children. *Zhongguo Dang Dai Er Ke Za Zhi*. 11(9): 769-770.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. New York: Oxford University Press.

Dowling, A.F. (1980) Do hospital staff interfere with computer system implementation? *Healthcare Management Review* 5:23-32.

Draper, M., Cohen, P., Buchan, H. (2001) Seeking consumer views: what use are results of hospital patient satisfaction surveys? *International Journal for Quality in Health Care* 13(6):463 –8.

Duggan, P.F. (1992) Time to abolish "gold standard." *British Medical Journal* 304:1568-9.

Duncan, B., Barton, L., Edmonds, D., Blashill, B.M. (2004). Parental perceptions of the therapeutic effect from osteopathic manipulation or acupuncture in children with spastic cerebral palsy. *Clin Pediatr* (Phila). 43(4): 349-353.

Duncan, E.A., Murray, J. (2012) The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC Health Services Research*12:96.

Dunn, A.S., Green, B.N., Formolo, L.R., Chicoine, D. (2011) Retrospective case series of clinical outcomes associated with chiropractic management for veterans with low back pain. *J Rehabil Res Dev*. 48(8):927-34.

Dunn, K.M., Campbell, P., Jordan, K.P. (2013) Long term trajectories of back pain: cohort study with 7 year follow up. *BMJ Open* 3:e003838.

Dupeyron, A., Vautravers, P., Lecocq, J., Isner-Horobeti, M.E. (2003). Complications following vertebral manipulation-a survey of a French region physicians. *Ann Readapt Med Phys*. 46(1): 33-40.

Dusenbury, L., Brannigan, R., Falco, M., Hansen, W. (2003) A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research* 18:237-256.

Dworkin, R.H., Turk, D.C., Farrar, J.T. (2005) Core outcome measurements for chronic pain clinical trials: IMMPACT recommendations. *Pain* 113:9-19.

Dworkin, R.H., Turk, D.C., Wyrwich, K.W. (2008) Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *Journal of Pain* 9:105-21.

Ebrall, P. S. (1994). Some anthropometric dimensions of male adolescents with idiopathic low back pain. *J Manipulative Physiol Ther*. 17(5): 296-301.

Eckert, J.P. Jr., Mauchly, J.W. (1947) *Electronic Numerical Integrator and Computer*, United States Patent Office, US Patent 3,120,606, filed 26 June 1947, issued 4 February 1964, and invalidated 19 October 1973 after court ruling on Honeywell v. Sperry Rand. *Retrieved 21st July, 2014 from* https://en.wikipedia.org/wiki/History_of_computing_hardware.

Eddy, D.M. (1982a) Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities. In Kahneman, D.; Slovic, P.; Tversky, A. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Eddy, D.M. (1982b). Clinical Policies and the Quality of Clinical Practice. *New England Journal of Medicine* 307 (6): 343–7.

Eddy, D.M. (1984). Variations in Physician Practice The Role of Uncertainty. *Health Affairs* 3 (2): 74–89.

Eddy, D.M. (1988) The Quality of Medical Evidence: Implications for Quality of Care. *Health Affairs* 7 (1): 19–32.

Eddy, D.M. (1990) Practice Policies: Guidelines for Methods*. Journal of the American Medical Association* 263 (13): 1839–41

Edie, D. (1968). *Introduction to the Basic Computer. New York: Prentice-Hall.*

Edmondson, A.C. (2004) Learning from failure in health care: frequent opportunities, pervasive barriers. *Quality and Safety in Health Care* 13:ii3-ii9.

Edwards, A., Thomas, R., Williams, R., Ellner, A.L., Brown, P., Elwyn, G. (2006) Presenting risk information to people with diabetes: evaluating effects and preferences for different formats by a web-based randomised controlled trial. *Patient Education and Counselling* 63:336-49.

Edwards, P., Roberts, I., Clark, M. (2002) Increasing response rate to postal questionnaires: a systematic review. *British Medical Journal* 324:1183-5.

Edwards, T.B., Bostick, R.D., Greene, C.C. (2002) Interobserver and intraobserver reliability of the measurement of shoulder internal rotation by vertebral level. *Journal of Shoulder and Elbow Surgery* 11:40-42.

Effective Healthcare: *Implementation of clinical practice guidelines*. (1994). Bulletin Number 8. University of Leeds: Nuffield Institute for Health.

Ekblom, A., Hansson, P. (1988) Pain intensity measurements in patients with acute pain receiving afferent stimulation. *J Neurol Neurosurg Psychiatry.* 51(4):481-6.

Ekman, M., Johnell, O., Lidgren, L. (2005). The economic cost of low back pain in Sweden. Acta Orthop 76(2):275-284.

Eldridge, S.M., Chan, C.L., Campbell, M.J., Bond, C.M., Hopewell, S., Thabane, L., Lancaster, G.A.; PAFS consensus group. (2016) CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ.* 355:i5239.

El-Hawary, R., Karol, L.A., Jeans, K.A., Richards, B.S. (2008). Gait analysis of children treated for clubfoot with physical therapy or the Ponseti cast technique. *J Bone Joint Surg Am*. 90(7): 1508-1516.

Ellershaw, J., Gambles, M., McGlinchey, T. (2008). Benchmarking: a useful tool for informing and improving care of the dying? *Supportive Care in Cancer*, 16:813-819.

Elliott, D.S., Mihalic, S. (2004) Issues in disseminating and replicating effective prevention programs. *Prev Sci.* 5(1):47-53.

Ellis, B.M., Fitzpatrick, R., Hill, J.C., Price, A. (2014) Bridging the Musculoskeletal Measurement Gap. *Journal of Trauma and Orthopaedics* 2(4):24-25.

Ellis, J. (2006). Benchmarking in mental health: an introduction for psychiatrists. *Advances in Psychiatric Treatment*, 11:305-14.

Ellwood, P.M. (1988) Shattuck Lecture – outcomes management: a technology of patient experience. *New England Journal of Medicine* 318:1549-1556.

Emerson, R.M. (1976). Social Exchange Theory. *Annual Review of Sociology 2:335-362*

Emmelkamp, P.M. (2004) The additional value of clinimetrics needs to be established rather than assumed. *Psychotherapy and Psychosomatics* 73(3):142-4.

Engel, G.L. (1977) The need for a new medical model: a challenge for biomedicine. *Science* 196(4286):129-136.

Engel, G.L. (1978) The biopsychosocial model and the education of health professionals. *Annals of the New York Academy of Sciences* 310:169-181.

Engels, Y., Dautzenberg, M., Campbell, S. (2006) Testing a European set of indicators for the evaluation of the management of primary care practices. *Fam Pract* 23:137-147.

Epstein, R.S. (2000) Responsiveness in quality of life assessment, nomenclature, determinants, and clinical applications. *Medical Care* 38:1191-94.

Erickson, K., et al. (2006). Case study in integrative medicine: Jared C, a child with recurrent otitis media and upper respiratory illness. *Explore* (NY). 2(3): 235-237.

Ernst, E., Pittler, M.H. (1999) Experts' opinions on complementary/alternative therapies for low back pain. *Journal of Manipulative & Physiological Therapeutics* 22(2):87-90.

Espallargues, M., Valderas, J.M., Alonso, J. (2000). Provision of feedback on feedback on perceived health status to healthcare professionals: a systematic review of its impact. *Medical Care*, 38:175-186.

Etikan, I., Musa, S.A., Alkassim, R.S. (2016) Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics* 5(1): 1-4

Ettorchi-Tardivey, A., Levie, M., Michel, P. (2012). Benchmarking: a method for continuous quality improvement in health. *Healthcare Policy*, 7(4):e101-119).

European Commission. *Patient Involvement*. Retrieved 20th February, 2016 from http://ec.europa.eu/health/systems_performance_assessment/docs/eurobaro_patient_involvement_2012_en.pdf.

European Low Back Pain Working Group (ELBPWG). (2004a) *European Guidelines for the Management of acute non-specific low back pain in primary care*. Retrieved 6th May, 2014 from www.backpaineurope.org/web/html/wg1_results.html.

European Low Back Pain Working Group (ELBPWG). (2014b) *European Guidelines for the Management of chronic non-specific low back pain*. Retrieved 6th May, 2014 from www.backpaineurope.org/web/html/wg2_results.html.

EuroQol Group. (1990) EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 16:199–208.

Evans, D.W. (2002) Mechanisms and effects of spinal high-velocity, low-amplitude thrust manipulation: previous theories. *Journal of Manipulative and Physiological Therapeutics* 25(4):251-62.

Evans, D. (2003) Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. Journal of Clinical Nursing. 12(1):77

Evans, D.W., Breen, A.C. (2006) A biomechanical model for mechanically efficient cavitation production during spinal manipulation: prethrust position and the neutral zone. *Journal of Manipulative and Physiological Therapeutics* 29(1):72-82.

Evidence-Based Medicine Working Group (1992) Evidence-based medicine. A new approach to teaching the practice of medicine. *Journal of the American Medical Association* 268 (17): 2420–5.

Fairbank, J.C., Couper, J., Davies, J.B., O'Brien, J.P. (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271-273.

Fanurik, D., Koh, J.L., Harrison, R.D., Conrad, D.M., Tomerlin, C. (1998) Pain assessment in children with cognitive impairment. An exploration of self-report skills. *Clinical Nursing Research* 7;103-19.

Fargnoli, V., Petignat, P., Burton-Jeangros, C. (2015) To what extent will women accept HPV self-sampling for cervical cancer screening? A qualitative study conducted in Switzerland. *International Journal of Women's Health* 7:883-8.

Farrar, J.T., Portnenoy, R.K., Berlin, J.A., Kinman, J.L., Strom, B.L. (2001) Defining the clinically important differences in pain outcome measures. *Pain*, 88:287-294.

Fasolo, B., Reutskaja, E., Dixon, A., Boyce, T. (2010) Helping patients choose: how to improve the design of scorecards of hospital quality. *Patient Education and Counselling* 78:344-9.

Fassaert, T., van Dulmen, S., Schelleris, F. (2008) Raising positive expectations helps patients with minor ailments: A cross-sectional study. *BMC Family Practice.* 9:38.

Fava, G.A., Tomba, E., Sonino, N. (2012) Clinimetrics: the science of clinical measurements. *International Journal of Clinical Practice* 66(1):11-5.

Fawkes, C.A. (2007) *Patient expectation and satisfaction with osteopathy in private practice*. Masters Thesis. London, Barts and The London Queen Mary's School of Medicine and Dentistry.

Fawkes, C.A., Leach, C.M., Mathias, S., Moore, A.P. (2014a) Development of a data collection tool to profile osteopathic practice: use of a nominal group technique to enhance clinician involvement. *Manual Therapy* 19(2):119-24.

Fawkes, C.A., Leach, C.M., Mathias, S., Moore, A.P. (2014b) A profile of osteopathic care in private practices in the United Kingdom: a national pilot using standardised data collection. *Manual Therapy* 19(2):125-30.

Fayaz, A., Croft, P., Langford, R.M., Donaldson, L.J., Jones, G.T. (2016) Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. *British Medical Journal Open* 6(6):e010364.

Feinstein, A.R. (1967). *Clinical Judgement*. Baltimore: Williams & Wilkins.

Feinstein, A. (1982) T Duckett Jones Memorial Lecture. The Jones criteria and the challenge of clinimetrics. *Circulation* 66:1-5.

Feinstein, A.R. (1987) *Clinimetrics*. Westford, MA: Yale University.

Fejer, R., Ruhe, A. (2012). What is the prevalence of musculoskeletal problems in the elderly population in developed countries? A systematic critical literature review. *Chiropractic and Manual Therapy* 20:31.

Fenton, J.J., Jerant, A.F., Bertakis, K.D., Franks, P. (2012). The cost of satisfaction: a national study of patient satisfaction, healthcare utilisation, expenditures, and mortality. *Archives of Internal Medicine* 172:405-41.

Ferraz, M.B., Quaresma, M.R., Aquino, L.R., Atra, E., Tugwell, P., Goldsmith, C.H. (1990). Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *Journal of Rheumatology* 17: 1022–4.

Festinger, L. (1954) A history of social comparison processes. *Hum Relat* 7:117-140.
Field, J.R. (2014). Personal correspondence concerning developments in PROMs in manual therapies.

Field, J.R., Newell, D. (2016) Clinical Outcomes in a Large Cohort of Musculoskeletal Patients Undergoing Chiropractic Care in the United Kingdom: A Comparison of Self- and National Health Service-Referred Routes. *Journal of Manipulative and Physiological Therapeutics* 39(1):54-62.

Fine, L.H., *The SWOT Analysis: Using your Strength to Overcome Weaknesses, Using Opportunities to Overcome Threats,* Kickit LLC, 2010, London.

Finlay, L. (2002) "Outing" the researcher: The provenance, process, and practice of reflexivity. *Qualitative Health Research* 12: 531–545.

Finocchietti, S., Graven-Nielsen, T., Arendt-Nielsen, L. (2015) Dynamic mechanical assessment of muscle hyperalgesia in humans: the dynamic algometer. *Pain Research and Management* 20(1):29-34.

Fishbein, M., Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley.

Fitzpatrick, R. (1992) Quality of life measures in healthcare. I: applications and issues in assessment. *British Medical Journal* 305:1074-1077.

Fitzpatrick, R., Davey, C., Buxton, M.J. (1998) Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 2(14):1-74).

Fontana, A., Frey, J.H. (2000) The interview from structured questions to negotiated text In N.K. Denzin and Y.S. Lincoln (eds) *Handbook of Qualitative Research*, 2nd edition, Thousand Oaks, CA: Sage.

Forbat, L., Cayless, S., Knighting, K. (2009) Engaging patients in health care: an empirical study of the role of engagement on attitudes and action. *Patient Education and Counselling* 74:84–90.

Ford, T., Hutchings, J., Bywater, T., Goodman, A., Goodman, R. (2009). Strengths and Difficulties Questionnaire added value scores: Evaluating effectiveness in child mental health interventions. *British Journal of Psychiatry*, 194(6), 552–558.

Fox, E.J., Melzack, R. (1976) Transcutaneous nerve stimulation and acupuncture: comparison of treatment for low back pain. *Pain* 2:141-8.

Francis, D.O., McPheeters, M.L., Noud, M., Penson, D.F., Feurer, I.D. (2016) Checklist to operationalize measurement characteristics of patient-reported outcome measures. Systematic Reviews 5(1):129.

Francis, J.J., Tinmouth, A., Stanworth, S.J., Grimshaw, J.M., Johnston, M., Hyde, C., Stockton, C., Brehaut, J.C., Fergusson, D., Eccles, M.P. (2009) Using theories of behaviour to understand transfusion prescribing in three clinical contexts in two countries: development work for an implementation trial. *Implement Sci.* 4:70.

Francis, R. Report of the enquiry into the Mid Staffordshire NHS Trust, 2013. Retrieved 13th May, 2014 from http://www.midstaffspublicinquiry.com/report.

Frank, A.J., Moll, J.M., Hort, J.F. (1982) A comparison of three ways of measuring pain. *Rheumatology and Rehabilitation*. 21:211-217.

French, K., Diamond, E., Gronkiewicz, C., Borkgren, M. (2010) Electronic medical records: a practitioner's perspective on evaluation and implementation. *Chest.* 138(3):716-23.

Freyd, M. (1923) The graphic rating scale. *Journal of Educational Psychology* 14:83-102.

Fries, J.F., Cella, D., Rose, M., Krishnan, E., Bruce, B. (2009) Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. Journal of Rheumatology 36(9):2061–2066.

Frost, H., Lamb, S.E., Stewart-Brown, S. (2008) Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index v2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. *Spine* 33(22):2450-2457.

Froud, R., Abel, G. (2014) Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. theoretical considerations and an example application of change in health status. *PLoS One*. 9(12):e114468.

Froud, R., Rajendran, D., Fossum, C. (2008) How do patients feel post-treatment? Pilot study at a UK osteopathic teaching clinic of self-reported adverse events. *International Journal of Osteopathic Medicine* 11(4):151-152.

Froud, R., Eldridge, S., Lall, R., Underwood, M. (2009) Estimating the number needed to treat from continuous outcomes in randomised controlled trials: methodological challenges and worked example using data from the UK Back Pain Exercise and Manipulation (BEAM) trial. *BMC Medical Research Methodology* 9:35.

Froud, R., Ellard, D., Patel, S., Eldridge, S., Underwood, M. (2015) Primary outcome measures used in back pain trials may need radical reassessment. *BMC Musculoskeletal Disorders* 16:88.

Froud, R., Patterson, S., Eldridge, S., Seale, C., Pincus, T., Rajendran, D. (2014) A systematic review and meta-synthesis of the impact of low back pain on people's lives. *BMC Musculoskeletal Disorders* 15:50.

Fu, Y., McNichol, E., Marczewski, K., Closs, S.J. (2016) Exploring the Influence of Patient-Professional Partnerships on the Self-Management of Chronic Back Pain: A Qualitative Study. *Pain Management and Nursing* S1524-9042(16)30040-6.

Furlan, A.D., Imamura, M., Dryden, T. (2009) Massage for low back pain: an updated systematic review within the framework of the Cochrane Back Review Group. *Spine* 34(16):1669-84.

Furnham, A., Vincent, C., Wood, R. (1995) The health beliefs and behaviours of three groups of complementary medicine and a general practice group of patients. *Journal of Alternative and Complementary Medicine.* 1(4): 347-59.

Gallagher, E.J., Bijur, P.E., Latimer, C., Silver, W. (2002) Reliability and validity of a visual analogue scale for acute abdominal pain in the ED. *American Journal of Emergency Medicine* 20:287-290.

Garbett, R., McCormack, B. (2001) The experience of practice development: An exploratory telephone interview study. *Journal of Clinical Nursing* 10:94–102.

Garratt, A., Schmidt, L., Mackintosh, A., Fitzpatrick R. (2002) Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal* 324(7351):1417.

Gartner (2016) Worldwide Devices Shipments by Device Type, 2015-2018 (Millions of Units). Retrieved 21st July, 2016 from http://www.gartner.com/newsroom/id/3187134.

Gault, R.H. (1907) A history of the questionnaire method of research in psychology. *Research in Psychology* 14 (3): 366–383.

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. *J Manipulative Physiol Ther*. 30(4):259-62.

Gemmell, H., Miller, P. (2010) Relative effectiveness and adverse effects of cervical manipulation, mobilisation and the activator instrument in patients with sub-acute non-specific neck pain: results from a stopped randomised trial. *Chiropr Osteopat.* 9;18:20.

General Chiropractic Council (GCC). (2015) *GCC Annual Review 2015*. Retrieved 20th April, 2016 from http://www.gcc-uk.org/UserFiles/Docs/Annual%20Report/Annual%20Review%20Final.pdf.

General Osteopathic Council (1998) *Osteopathy Snapshot Survey*. Unpublished results. London: General Osteopathic Council.

General Osteopathic Council (2001). *Snapshot Survey 2001.* Unpublished results. London: General Osteopathic Council. Retrieved 20th April, 2016 from http://www.osteopathy.org.uk/uploads/survey2snapshot_survery_results_2001.pdf.

General Osteopathic Council (2012). *Osteopathic Practice Standards*. General Osteopathic Council. Retrieved 20th April, 2016 from http://www.osteopathy.org.uk/uploads/osteopathic_practice_standards_public.pdf.

General Osteopathic Council (2015a) GOsC-ASA-CAP guidance for osteopaths. Retrieved 16th September, 2015 from http://www.osteopathy.org.uk/news-and-resources/document-library/practice-guidance/gosc-asa-cap-guidance-for-osteopaths/.

General Osteopathic Council (2015b). *Registration Report to Council, May, 2015*. Retrieved 20th April, 2016 from http://www.osteopathy.org.uk/training-and-registration/.

General Osteopathic Council (2015c). *Public Perceptions Study full report, June 2015*. Retrieved 20th April, 2016 from http://www.osteopathy.org.uk/news-and-

resources/document-library/research-and-surveys/public-perceptions-study-full-report/ .

General Osteopathic Council (2016a). *(GOsC) Registration Report, May, 2016*. Retrieved 20th April, 2016 from http://www.osteopathy.org.uk/news-and-resources/document-library/about-the-gosc/council-may-2016-item-16-

registration-report/.

General Osteopathic Council (2016b). *Number of osteopaths registered with the GOsC, August, 2016*. Retrieved 20th April, 2016 from http://www.osteopathy.org.uk/news-and-resources/research-surveys/statistics/ .

GOsC, 2017. New CPD requirements from 2018 onwards. http://cpd.osteopathy.org.uk/resource/patient-reported-outcome-measures/.

General Practitioners' Committee (2002) *Your contract, your future*. London: BMA.

Georgy, E.E., Carr, E.C., Breen, A.C. (2013) Met or matched expectations: what accounts for a successful back pain consultation in primary care? *Health Expectations*. 16: 143-54.

Geri, T., Signori, A., Gianola, S., Rossettini, G., Grenat, G., Checchia, G., Testa. M. (2015) Cross-cultural adaptation and validation of the Neck Bournemouth Questionnaire in the Italian population. *Qual Life Res.* 24(3):735-45.

Gerteis, M., Gerteis, J.S., Newman, D., Koepke, C. (2007) Testing consumers' comprehension of quality measures using alternative reporting formats. *Health Care Finance Review* 28:31-45.

Gibbons, P., Tehan, P. (2006) HVLA thrust techniques: What are the risks? *International Journal of Osteopathic Medicine* 9(1): 4-12.

Gibson, T., Grahame, R., Harkness, J. (1985) Controlled comparison of short-wave diathermy treatment with osteopathic treatment in non-specific low back pain. *Lancet* 1:1258-1261.

Giesen, J. M., Center, D.B., Leach, R.A. (1989). An evaluation of chiropractic manipulation as a treatment of hyperactivity in children. *J Manipulative Physiol Ther*. 12(5): 353-363.

Gilbody, S.M., House, A.O., Sheldon, T.A. (2001). Routine administered questionnaires for depression and anxiety: systematic review. *British Medical Journal,* 322:406-409.

Gillespie, B. R. (2009). Case study in attention-deficit/hyperactivity disorder: the corrective aspect of craniosacral fascial therapy. *Explore* (NY). 5(5): 296-298.

Gillies, R.R., Shortellm S,M., Casalino, L. (2003) How different is California? A comparison of US physician organizations. *Health Affairs (Millwood)*. Retrieved 29th May, 2015 from http://content.healthaffairs.org./cgi/reprint/hlthaff.w3.492v1.pdf.

Glaser, B., Strauss, A. (1967) *The discovery of grounded theory: strategies for qualitative research*. Chicago: Aldine.

Glattacker, M., Heyduck, K., Meffert, C. (2013). Illness beliefs and treatment beliefs as predictors of short-term and medium-term outcome in chronic back pain. *Journal of Rehabilitation Medicine*, 45:268-276.

Godin, G., Bélanger-Gravel, A., Eccles, M., Grimshaw, J. (2008) Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. *Implement Sci*. 3:36.

Goertz, C.M., Pohlman, K.A., Vining, R.D., Brantingham, J.W., Long, C.R. (2012). Patient-centered outcomes of high-velocity, low-amplitude spinal manipulation for low back pain: a systematic review. *Journal of Electromyography and Kinesiology* 22(5):670-91.

Goeschel, C.A., Weiss, W.M., Pronovost, P.J. (2012) . Using a logic model to design and evaluate quality and patient safety improvement programs. *International Journal for Quality in Health Care* 24(4):330-337.

Gottesman, W., Baum, N. (2013) QR Codes: Next Level of Social Media. *The Journal of Medical Practice Management.* 345-7.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25:161-178.

Gracely, R.H., Kwilosz, D.M. (1988). The Descriptor Differential Scale: applying psychophysical principles to clinical pain assessment. *Pain*. 35 (3): 279–88

Graz, B., Wiretlisbach, V., Porchet, F. (2005) Prognosis or "curabo effect"? Physician prediction and patient outcome of surgery for low back pain and sciatica. *Spine.* 30(12): 1448-52.

Green, M.F., Kern, R.S., Heaton, R.K. (2004) Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. Schizophr Res. 72(1):41-51.

Greenhalgh, J., Meadows, K. (1999). The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of care: a literature review. *Journal of Evaluation in Clinical Practice*, 5:401-406.

Greenhalgh, J., Long, A.F., Flynn, R. (2005) The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? Soc Sci Med. 60(4):833-43.

Greenhalgh, J. (2009) The application of PROs in clinical practice: what are they, do they work, and why? *Quality of Life Research* 18:115-123.

Greenhalgh J, Dalkin S, Gooding K, Gibbons E, Wright J, Meads D, Black N, Valderas JM, Pawson R (eds). (2017) Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Southampton (UK): NIHR Journals Library.

Greenman, PE. (1989) *Principles of Manual Medicine*. Baltimore, Md: Williams and Wilkins.

Greenough, C.G. (2006). Outcome assessment: recommendations for daily practice. *European Spine Journal*, 15:S118-S123.

Greenough, C. (2017). National Back Pain Pathway. http://www.ukssb.com/pages/Improving-Spinal-Care-Project/National-Backpain-Pathway.html.

Greenwood, M.C., Hakim, A.J., Carson, E. (2006) Touch-screen computer systems in the rheumatology clinic offer a reliable and user-friendly means of collecting quality-of-life and outcome data from patients with rheumatoid arthritis. *Rheumatology* 45:66 –71

Grégoire, G., Derderian, F., Le Lorier, J. (1995). Selecting the language of the publications included in a meta-analysis: Is there a tower of babel bias*?* *Journal of Clinical Epidemiology* 48(1):159-163.

Greiffenhagen, C., Mair, M., Sharrock, W. (2015) Methodological troubles as problems and phenomena: ethnomethodology and the question of 'method' in the social sciences. *British Journal of Sociology* 66(3):460-85.

Griffin, B., Baston, L. (2014) June 2014. The user interface of a mechanical system, a vehicle or an industrial installation is sometimes referred to as the human-machine interface (HMI). *Interfaces* 5(7)

Griffiths, R. (1983) *NHS Management Inquiry Report*. DHSS, London.

Grimmer, K., Sheppard, L., Pitt, M. (1999) Differences in stakeholder expectations in the outcome of physiotherapy management of acute low back pain. *International Journal for Quality in Health Care* 11(2): 155-62.

Grimshaw, J., Freemantle, N., Wallace, S., Russell, I., Hurwitz, B., Watt, I., Long, A., Sheldon, T. (1995) Developing and implementing clinical practice guidelines. *Qual Health Care*. 4(1):55-64.

Grol, R. (1992)Implementing guidelines in general practice care*. BMJ.* 315:418-421.

Grol, R. (1997) Personal paper. Beliefs and evidence in changing clinical practice. *BMJ*. 315(7105):418-21.

Grol, R., Grimshaw, J. (2003) From Best Evidence to Best Practice: Effective Implementation of Change in Patients' Care. *Lancet.* 11(362):1125–30.

Grol, R., Wensing, M. (2004) What Drives Change? Barriers to and Incentives for Achieving Evidence-Based Practice. *Medical Journal of Australia*. 180(6):S57–S60.

Grol, R., Wensing, M., Eccles, M., Davis, D. (2013*) Improving patient care: the implementation of change in healthcare.* London: BMJ Books, Wiley Blackwell.

Grotle, M., Brox, J.I., Vøllestad, N.K. (2004) Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine* 29(21):E492-501.

Groves, R. (1990) Theories and methods of telephone surveys. *Annual Review of Sociology* 16:221– 240.

Grøvle, L., Haugen, A.J., Hasvil, E., Natvig, B., Brox, J.I., Grotle, M. (2014) Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. *Journal of Clinical Epidemiology* 67:508-515.

Guadagnioli, E., Ward, P. (1998) Patient participation in decision-making. *Social Science in Medicine* 47:329-39.

Guba, E.G. (1981). Criteria for assessing the trustworthiness of naturalistic enquiries. *Educational Communication and Technology Journal*, 29:75-91.

Guest, G., Bunce, A., Johnson, L. (2006). How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods* 18(1):59-82

Guo, J.B., Zhu, Y., Chen, B.L., Xie, B., Zhang, W.Y., Yang, Y.J., Yue, Y.S., Wang, X.Q. (2015). Surgical versus non-surgical treatment for vertebral compression fracture with osteopenia: a systematic review and meta-analysis. *PLoS One*. 10(5):e0127145.

Gurden, M., Morelli, M., Sharp, G., Baker, K., Betts, N., Bolton, J. (2012) Evaluation of a general practitioner referral service for manual treatment of back and neck pain. *Prim Health Care Res Dev*. 13(3):204-10.

Gurry, B., Hopkins, M., Peers, C., Anderson, S., Watts, M. (2004) A rapid access treatment facility for acute low back pain based in the primary care setting. *Journal of Orthopaedic Medicine* 26(1):13-19.

Guyatt, G.H., Juniper, E.F., Walter, S.D., Griffith, L.E., Goldstein, R.S. (1998) Interpreting treatment effects in randomised trials. *British Medical Journal* 316:690-693.

Guyatt, G.H, Norman, G.R., Juniper, E.F., Griffith, L.E. (2002a) A critical look at transition ratings. *Journal of Clinical Epidemiology* 55(9):900-8.

Guyatt, G.H., Osoba, D., Wu, A.W., Wyrwich, K.W., Norman, G.R. (2002b) Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings* 77: 371-383.

Gwaltney, C.J., Shields, A.L., Shiffman, S. (2008) Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-anaytical review. *Value in Health* 11(2):322-333.

Haefeli, M., Elfering, A. (2006) Pain assessment. *European Spine Journal* 15:S17-S24.
Hagen, E.M., Grasdal, A., Eriksen, H.R. (2003). Does early intervention with a light mobilization program reduce long-term sick leave for low back pain: A 3-year follow-up study. *Spine*. 28:2309–2316.

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

Hakkinen, A., Hannonen, P., Nyman, K., Lyyski, T., Hakkinen, K. (2003). Effects of concurrent strength and endurance training in women with early or longstanding rheumatoid arthritis: Comparison with healthy subjects. *Arthritis and Rheumatism* 49:789–797.

Halacy, D.S. (1970). *Charles Babbage, Father of the Computer.* University of Michigan: Crowell-Collier Press.

Hammersley, M., Atkinson, P. *Ethnography: principles in practice.* 2nd ed. London: Routledge.

Handberg, C., Thorne, S., Midtgaard, J., Nielsen, C.V., Lomborg, K. (2015) Revisiting Symbolic Interactionism as a Theoretical Framework Beyond the Grounded Theory Tradition. *Qualitative Health Research* 25(8):1023 –1032.

Hansen, A. B., Price, K.S., Loi, E.C., Buysse, C.A., Jaramillo, T.M., Pico, E.L., Feldman, H.M. (2014). Gait changes following myofascial structural integration (Rolfing) observed in 2 children with cerebral palsy. *J Evid Based Complementary Altern Med*. 19(4): 297-300.

Harris, D.A. (2015) Doing research drawing on the philosophy of existential hermeneutic phenomenology. *Palliative and Supportive Care* :1-3.

Harris, K.K., Dawson, J., Jones, L.D., Beard, D.J., Price, A.J. (2013) Extending the use of PROMs in the NHS – using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: a validation study. *British Medical Journal Open* 3:e003365.

Hart, C. (2003). *Doing a Literature Search A Comprehensive Guide for the Social Sciences*. London: Sage.

Hartvigsen, J., Lauridsen, H., Ekström, S. (2005) Translation and Validation of the Danish Version of the Bournemouth Questionnaire. *Journal of Manipulative and Physiological Therapeutics* 28(6):402–407.

Harvard University Behavioural models, Harvard University (2016). Retrived 12th February, 2016 from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/SB/SB721-Models/SB721-Models3.html.

Harvey, G., Oliver, K., Humphreys, J., Rothwell, K., Hegarty, J. (2014) Improving the identification and management of chronic kidney disease in primary care: lessons from a staged improvement collaborative. *Int J Qual Health Care.* 27(1):10-6.

Harvey, G., Kitson, A. (2015) Translating evidence into healthcare policy and practice: Single versus multi-faceted implementation strategies - is there a simple answer to a complex question? *Int J Health Policy Manag*. 5;4(3):123-6.

Haspeslagh, S.R.S., Van Suijlekom, H.A., Lamé, I.E., Kessels, A.G.H., van Kleef, M., Weber, W.E.J. (2006) Randomised controlled trial of cervical radiofrequency lesions as a treatment for cervicogenic headache. *BMC Anaesthesiology* 6:1.

Hatfield, D.R., Ogles, B.M. (2007) Why some clinicians use outcome measures and others do not. *Administration and Policy in Mental Health* 34(3):283–291.

Hawker, G.A., Mian, S., Kendzerska, T., French, M. (2011). Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care and Research* 63(s11): 240-252.

Hawthorne, K., Prout, H., Kinnersley, P., Houston, H. (2009) Evaluation of different delivery modes of an interactive e-learning programme for teaching cultural diversity. *Patient Education and Counselling* 74(1):5-11.

Hay, E.M., Dunn, K.M., Hill, J.C., Lewis, M., Mason, E.E., Konstantinou, K., Sowden, G., Somerville, S., Vohora, K., Whitehurst, D., Main, C.J. (2008) A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. *BMC Musculoskeletal Disorders*, 22;9:58.

Hayden, C., Mullinger, B. (2006) A preliminary assessment of the impact of cranial osteopathy for the relief of infantile colic. *Complement Ther Clin Pract*. 12(2):83-90.

Hayden, J. A., Mior, S.A., Verhoef, M.J. (2003). Evaluation of chiropractic management of pediatric patients with low back pain: a prospective cohort study. *J Manipulative Physiol Ther.* 26(1): 1-8.

Haydon, N.B. (2013) Head injury: audit of a clinical guideline to justify head CT. *Journal of Medical Imaging and Radiation Oncology* 57(2):161-8.

Hayes, N. M. Bezilla, T.A. (2006). Incidence of iatrogenesis associated with osteopathic manipulative treatment of pediatric patients. *J Am Osteopath Assoc*. 106(10): 605-608.

Hayes, H., Parchman, M.L., Howard, R. (2011). A logic model framework for evaluation and planning in a primary care practice-based research network (PBRN). *Journal of the American Board of Family Medicine* 24(5):576-582.

Hayes, M.J.H., Patterson, D.G. (1921) Experimental development of the graphic rating method. *Psychology Bulletin* 18:98.

Health and Social Care Act (2012). Retrieved 1st June, 2014 from http://www.legislation.gov.uk/ukpga/2012/7/contents/enacted.

Health and Social Care Act (2012) C7, Explanatory Notes. Retrieved 16th July, 2017 from http://www.legislation.gov.uk/ukpga/2012/7/notes.

Health and Social Care Information Centre (HSCIC) Retrieved 1st June, 2014 from http://www.hscic.gov.uk/.

Health and Social Care Information Service (HSCIC) (2015). *Provisional Monthly Patient Reported Outcome Measures (PROMs) in England – April 2014 to January 2015*. Retrieved 1st June, 2014 from http://www.hscic.gov.uk/catalogue/PUB17618.

Health Council of the Netherlands. (2000) From implementation to learning. The importance of two-way traffic between practice and science in healthcare. 2000/18. https://www.gezondheidsraad.nl/sites/default/files/0018e.pdf

Health Research Authority (2005). *Research Governance Framework: national versions*. Retrieved 20th June, 2014 from http://www.hra.nhs.uk/resources/research-legislation-and-governance/research-governance-frameworks/.

Hegarty, A.M., McGrath, C., Hodgson, T.A., Porter, S.R. (2002) Patient-centred outcome measures in oral medicine: are they valid and reliable? Oral and Maxillofacial Surgery 31(6):670-74.

Heilig, D. (1961) The thrust technique. *Journal of the American Osteopathic Association* 81:244-248.

Heilig, D. (1986) The 1984 Thomas L Northup memorial address: osteopathic manipulative care in preventive medicine. *Journal of the American Osteopathic Association* 86:645-651.

Hekkert, K.D., Cihangir, S., Kleefstra, S.M., van dn Berg, B., Kool, R.B. (2009). Patient satisfaction revisited: a multi-level approach. *Social Science and Medicine* 69:68-75.

Herbert, J.J., Fritz, J.M., Koppenhaver, S.L., Thackeray, A., Kjaer, P. (2016). Predictors of clinical outcome following lumbar disc surgery: the value of historical, physical examination, and muscle function variables. *European Spine Journal*, 25(1):310-7.

Herold D. In Penguin Dictionary of Modern Humerous Quotations p.84, Fred Metcalf. Penguin Books, London (2001).

Herr, K.A., Mobily, P.R. (1993) Comparison of selected pain assessment tools for use with the elderly. *Appl Nurs Res*. 6(1):39-46.

Herr, K., Spratt, K.F., Garand, L., Li, L. (2007) Evaluation of the Iowa pain thermometer and other selected pain intensity scales in younger and older adult cohorts using controlled clinical pain: a preliminary study. *Pain Med*. 8(7):585-600.

Hibbard, J.H., Greene, J., Daniel, D. (2010). What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. *Medical Care Research Review* 67:275-293.

Higgins, J.P., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savovic, J., Schulz, K.F., Weeks, L., Sterne, J.A.; Cochrane Bias Methods Group.; Cochrane Statistical Methods Group. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 343:d5928.

Hildenhovi, H., Nojonen, K., Laippala, P. (2002) Measurement of outpatients' views of service quality in a Finnish university hospital. *Journal of Advanced Nursing* 38(1):59 –67.

Hildon, Z., Allwood, D., Black, N. (2012a) Making data more meaningful: Patients' views of the format and content of quality indicators comparing health providers. *Patient Education and Counselling* 88:298-304.

Hildon, Z., Allwood, D., Black, N. (2012) Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. *International Journal for Quality in Health Care* 24(1):55-64.

Hildon, Z., Neuburger, J., Allwood, D., van der Meulen, J., Black, N. (2012) Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. *BMC Health Services Research* 2:171-184.

Hillebrand, F., Friedhelm, J. (eds) (2010) *The Creation of Global Mobile Communication*. Chichester, UK: John Wiley.

Hill, J.C., Kang, S., Bendetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K.M., Hay, E., Rees, J., Beard, D., Glyn-Jones, S., Barker, K., Ellis, B., Fitzpatrick, R., Price, A. (2016) Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *British Medical Journal Open* 6:e012331.

Hiller, H. H., & DiLuzio, L. (2004). The participant and the research interview: Analysing a neglected dimension in research. *The Canadian Review of Sociology and Anthropology* 41,1-26

Hinchcliffe, K.P., Surrall, K.E., Dixon, J.S. (1985) Reproducibility of pain measurements in rheumatoid arthritis by patients using visual analogue scales. *Pharmaceutical Medicine* 1:99-103.

Hinkley, H., Drysdale, I. (1995) Audit of 1000 patients attending the clinic of the British College of Naturopathy and Osteopathy. *British Osteopathic Journal* 16:17-27.

Hinman, R.S., Bennell, K.L., Crossley, K.M., McConnell, J. (2003). Immediate effects of adhesive tape on pain and disability in individuals with knee osteoarthritis. *Rheumatology* 42:865–869.

Hjermstad, M., Fayers, P.M., Haugen, D.F., Caraceni, A., Hanks, G.W., Loge, J.H., Fainsinger, R., Aass, N., Kaasa, S. (2011) Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. *Journal of Pain and Symptom Management* 41(6):1073-1093.

HM Government. Data Protection Act (1998). Retrieved 6th January, 2016 from (http://legislation.gov.uk/ukpga/1998/29/contents).

Hobus, P. (1994) *Expertise of primary care physicians*. Thesis, Maastricht: University of Maastricht.

Hoehler, F.K., Tobis, J.S., Buerger, A.A. (1981) Spinal manipulation for low back pain. *Journal of the American Medical Association* 245:1835-1838.

Hohwü, L., Lyshol, Hm, Gissler, M., Jonsson, S.H., Petzold, M., Obel, C. (2013) Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. *Journal of Medical Internet Research* 15(8):e173.

Holloway, I., Wheeler, S. (2015). Qualitative research in nursing and healthcare. Wiley-Blackwell; Oxford.

Hopkins A. (1990). *Measuring the quality of medical care*. London: Royal College of Physicians.

Horng, S., Goss, F.R., Chen, R.S. (2012) Prospective Pilot Study of a Tablet Computer in an Emergency Department. *International Journal of Medical Informatics* 81(5): 314 –319.

Hospital Episode Statistics (HES) (2014). Retrieved 1st May, 2014 from http://www.hscic.gov.uk/hes.

Hospital Episode Statistics for mortality 3_character_procedure_diagnosis_tables_V3_050613. Retrieved 2nd June, 2015 from http://www.hscic.gov.uk/hesonsmortality.

House of Lords. (2000) *Select Committee on Science and Technology: Sixth Report: Complementary and Alternative Medicine*. London: The Stationary Office.

Howell, D., Molloy, S., Wilkinson, K., Green, K., Orchard, K., Wang, K., Liberty, J. (2015). Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Annals of Oncology,* 26:1846–1858.

Howie, J.G., Heaney, D.J., Maxwell, M. (1999) Quality at general practice consultations: cross-sectional survey. *British Medical Journal* 319(7212):738-43)

Howie, R.R., Hamilton, D.F. (2013). PROMs – How was it for you? *Journal of Trauma and Orthopaedics* 1(1):49-51.

Hox, J.J., de Leeuw, E.D. (1994). A comparison of non-response in mail, telephone, and face-to-face surveys. *Quality and Quantity* 28:329-344.

Hsu, C., Bluespruce, J., Sherman, K., Cherkin, D. (2010) Unanticipated benefits of CAM therapies for back pain: an exploration of patient experiences. *J Altern Complement Med*. 16(2):157-63.

Huber, A., Suman, A.L., Rendo, C.A., Biasi, G., Marcolongo, R., Carli, G. (2007) Dimensions of "unidimensional" ratings of pain and emotions in patients with chronic musculoskeletal pain. *Pain*. 130(3):216-24.

Huckvale, K., Prieto, J.T., Tilney, M., Benghozi, P-J., Car, J. (2015) Unaddressed privacy issues in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC Medicine*13:214.

Hudak, P.M., Wright, J.P. (2003) The metaphor of patients as customers: Implications for patient satisfaction. *Journal of Clinical Epidemiology* 56(2): 103-8.

Hudon, A., Drolet, M.J., Williams-Jones, B. (2015) Ethical issues raised by private practice physiotherapy are more diverse than first meets the eye: recommendations from a literature review. *Physiotherapy Canada* 67(2):124-32.

Hudson-Cook, N., Tomes-Nicholson, K., Breen, A.C. (1989) A revised Oswestry disability questionnaire. In: Roland M, Jenner J. editors. *Back pain. New approaches to rehabilitation and education.* Manchester: Manchester University Press.

Hughes, C.L. *Introduction to research methodologies*. Retrieved 12th May, 2016 from http://www2.warwick.ac.uk/fac/soc/sociology/staff/hughes/researchprocess/introduction_to_research_methodologies.docx .

Hughes, D.L., DuMont, K. (2002) Using Focus Groups to Facilitate Culturally Anchored Research in *Ecological Research to Promote Social Change*. Springer, USA.

Hulscher, M., Wensing, M., Grol, R. (2000) *Effective Implementation. Theories and Strategies* Report ZON, Den Haag.

Humphreys, B. K. (2010). Possible adverse events in children treated by manual therapy: a review. *Chiropr Osteopat*. 18:12.

Humphreys, B.K., Peterson, C. (2013) Comparison of outcomes in neck pain patients with and without dizziness undergoing chiropractic treatment: a prospective cohort study with 6 month follow-up. *Chiropr Man Therap*. 7;21(1):3.

Hunt, S.M., McEwen, J. (1980) The development to a subjective health indicator *Sociology, Health, and Illness* 2:231–246.

Hunt, J. (2013). *NHS challenged to go paperless by 2018*. Retrieved 21st July, 2016 from https://www.gov.uk/government/news/jeremy-hunt-challenges-nhs-to-go-paperless-by-2018--2).

Hunter, C. (2014). Personal communication concerning developments in PROMs use among clinicians. http://oxford.academia.edu/CherylHunter

Hunter, C., Fitzpatrick, R., Jenkinson, C., Darlington, A-S.E, Coulter, A., Forder, J.E., Peters, M. (2015) Perspectives from health, social care and policy stakeholders on the value of a single self-report outcome measure across long-term conditions: a qualitative study. *British Medical Journal Open* 5:e006986.

Hurst, H., Bolton, J. (2004) Assessing the clinical significance of change scores recorded on subjective outcome measures. *Journal of Manipulative and Physiological Therapeutics* 27(1):26-35.

Hush, J.M., Kamper, S.J., Stanton, D.R., Ostelo, R., Refshauge, K.M. (2012) Standardized measurement of recovery from non-specific back pain. *Archives of Physical Medicine and Rehabilitation* 93(5):849-55.

Hush, J.M., Refshauge, K.M., Sullivan, G., De Souza, L., McAuley, J.H. (2010) Do numerical rating scales and the Roland Morris Disability Questionnaire capture

changes that are meaningful to patients with persistent back pain? *Clinical Rehabilitation* 24(7):648-57.

Husted, J.A., Cook, R.J., Farewell, V.T., Gladman, D.D. (2000) Methods for assessing responsiveness: a critical review and recommendations. *Journal of Clinical Epidemiology* 53:459-468.

IASP (1986) *Classification of chronic pain descriptions of chronic pain syndromes and definitions of pain terms*. (2nd ed). Seattle: IASP Press.

Iles, R.A., Davidson, M., Taylor, N.F. (2008). Psychosocial predictors of failure to return to work in non-chronic non-specific low back pain: a systematic review. *Occupational and Environmental Medicine*, 65:507-517.

Indahl, A., Velund, L., Reikeraas, O. (1995). Good prognosis for low back pain when left untampered: a randomised clinical trial. *Spine*, 20:473-7.

Independent Information Governance Oversight Panel (IIGOP). Report on Care.Data, 2014. Retrieved 4th June, 2015 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/389219/IIGOP_care.data.pdf.

Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT). Retrieved 6th June, 2015 from http://www.immpact.org/.

International Consortium for Health Outcomes (ICHOM) https://www.ichom.org/. Isis Innovations. http://isis-innovation.com/health-outcomes/.

Ivar Brox, J., Sorensen, R., Friis, A., Nygaard, O., Indahl, A., Keller. A., Ingebrigtsen, T., Eriksen, H.R., Holm, I., Koller, A.K., Riise, R., Reikeras, O. (2003). Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 28:1913–1921.

Jacobson, N.S., Follette, W.G., Revenstorf, D. (1984) Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behavioural Therapy*15:336-52.

Jackson, J., Kroenke, K. (2001) The effect of unmet expectations among adults presenting with physical symptoms. *Annals of Internal Medicine*. 2001;134: 889-98.

Jackson, J.L., Kroenke, K., Chamberlin, J. (2001) Effects of physician awareness of symptoms-related expectations and mental disorders: A controlled trial. *Archives of Family Medicine*. 8(2):135-42.

Jadad, A.R., Enkin, M.W. (2007) *Randomized controlled clinical trials*, 2nd edition. London: BMJ Books.

Jaeschke, R., Singer, J., Guyatt, G.H. (1989) Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 10(4):407-15.

Jamison, R.N., Graceley, R.H., Raymond, S.A., Levine, J.G., Marino, B., Herrman, T.J., Daly, M., Fram, D., Katz, N.P. (2002) Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers. *Pain* 99:341-347.

Janwantanakul, P., Sihawong, R., Sitthipornvorakul, E., Paksaichol, A. (2015). A screening tool for non-specific low back pain with disability in office workers: a 1-year prospective cohort study. *BMC Musculoskeletal Disorders* 16: 298.

Jellema, P., van der Horst, H.E., Vlaeyan, J.W., Stalman, W.A., Bouter, L.M., van der Windt, D.A. (2006). Predictors of outcome in patients with (sub)acute low back pain differ across treatment groups. *Spine*, 31(15):1699-705.

Jenkins, P.J., Sng, S., Brooksbank, K., Brooksbank, A.J. (2016) Socioeconomic deprivation and age are barriers to the online collection of patient reported outcome measures in orthopaedic patients. *Annals of the Royal College of Surgeons of England* 98(1):40-4.

Jenkinson, C., McGee, H. (1998) *Health status measurement*. Oxford: Radcliffe.

Jensen, M.P., Karoly, P., Braver, S. (1986) The measurement of clinical pain intensity: a comparison of six methods. *Pain* 27:117-126.

Jensen, M.P., Karoly, P., O'Riordan, E.F., Bland, F., Burns, R.S. (1989) The subjective experience of acute pain. An assessment of the utility of 10 indices. *Clinical Journal of Pain* 5:153-9.

Jette, D.U., Halbert, J., Iverson, C., Miceli, E., Shah, P. (2009) Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Physical Therapy* 89(2):125–135.

Jobber, D., Ellis-Chadwick, F. (2012). *Principles and Practice of Marketing*. Boston; McGraw-Hill Education.

Johnson, R.B., Onwuegbuzie, A.J. (2004) Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher* 33(7):14–26.

Joint Commission on Accreditation of Healthcare Organizations (JCAHO, 2013). *Pain: current understanding of assessment, management and treatments.* Joint Commission on Accreditation of Healthcare Organizations and the National Pharmaceutical Council, Inc. December 2001. Retrieved 4th January, 2016.

Joyce, C.R., Zutshi, D.W., Hrubes, V., Mason, R.M. (1975) Comparisons of fixed interval and visual analogue scales for rating chronic pain. *European Journal of Clinical Pharmacology* 8:415-420.

Jones, D.S., Podolsky, S.H. (2015) The history and fate of the gold standard. *Lancet* 385(9977):1502-3.

Jones HL. (1981) *Strain and Counterstrain*. Newark, Ohio: American Academy of Osteopathy.

Joyce, C.R., Zutshi, D.W., Hrubes, V., Mason, R.M. (1975) Comparison of fixed interval and visual analogue scales for rating chronic pain. *European Journal of Clinical Pharmacology* 8:415-420.

Juniper, E.F., Guyatt, G.H., Wilan, A., Griffith, L.E. (1994) Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *Journal of Clinical Epidemiology* 47:81-87.

Kamper, S.J., Apeldoorn, A.T., Chiarotto, A., Smeets, R.J., Ostelo, R.W., Guzman, J., van Tulder, M.W. (2015). Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. *British Medical Journal* 350:h444.

Kaplan, B. (2001a) Evaluating informatics applications – clinical decision support systems literature review. *International Journal of Medical Informatics*. 64:15-37.

Kaplan, B. (2001b) Evaluating informatics applications – some alternative approaches: Theory, social interactionism, and call for methodological pluralism. *International Journal of Medical Informatics*. 64:39-56.

Kaplan, B., Shaw, N.T. (2004) People, organisational, and social issues: Future directions in evaluation research. *Methods of Information in Medicine* 43:215-231.

Kaplan, B., Maxwell, J.A. (2005) In Anderson JG, Aydin C. *Evaluating the Organisational Impact of Health Care*. New York: Springer.

Kaplowitz, M.D., Hadlock, T.D., Levine, R. (2004) A comparison of web and mail survey response rates. *Public Opinion Quarterly* 68(1):94-101.

Kaptchuk, T.J., Kelley, J.M., Conboy, L.A, (2008). Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *British Medical Journal.* 336(7651):999–1003.

Karstens, S., Hermann, K., Froböse, I., Weiler, S.W. (2013). Predictors for half-year outcome of impairment in daily life for back pain patients referred for physiotherapy: a prospective observational study. *PLOS one*, 8(4):e61587.

Katz, J., Wandersman, A., Goodman, R.M., Griffin, S., Wilson, D.K., Schillaci, M. (2013) Updating the FORECAST formative evaluation approach and some implications for ameliorating theory failure,implementation failure, and evaluation failure. *Eval Program Plann.* 39:42-50.

Kazis, L.E., Callahan, L.F., Meenan, R.F., Pincus, T. (1990). Health status reports in the care of patients with rheumatoid arthris. *Journal of Clinical Epidemiology*, 43:1243-1253.

Keck, J.F., Gerkensmeyer, J.E., Joyce, B.A., Schade, J.G. (1996) Reliability and validity of the Faces and Word Descriptor Scales to measure procedural pain. *Journal of Paediatric Nursing* 11:368-74.

Kelly, A. (2001). The minimum clinically significant difference in visual analogue scale pain score does not differ with severity of pain. *Journal of Emergency Medicine* 18(3):205-207.

Kelly, K.D., Voaklander, D.C., Johnston, D.W., Newman, S.C., Suarez-Alamazar, M.E. (2001). Change in pain and function while waiting for major joint arthroplasty. *Journal of Arthroplasty*, 16:351-9.

Kemshall, H., Littlechild, R. (eds) (2000). *User Involvement and Participation in Social Care: Research Informing Practice*, London: Jessica Kingsley Publishers.

Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., Bhattacharya, A., Gimson, A., Brütt, A.L., de Jong, K., Moore, M. (2016). Routine use of Patient Reported Outcome Measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*, 7, Art No.: CD011119.

Kennedy, D.M., Stratford, P.W., Wessel, J., Gollish, J.D., Penney, D. (2005). Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. *BMC Musculoskeletal Disorders*, 6:3-15.

Kennedy, T., Lingard, L. (2006) Making sense of grounded theory. *Medical Education* 40:101-8.

Kent, P., Grotle, M., Dunn, K.M., Albert, H.B., Lauridsen, H. (2015) Rasch analysis of the 23-item version of the Roland Morris Disability Questionnaire. *Journal of Rehabilitation Medicine*. 47:356-364.

Kerns, R.D., Turk, D.C., Rudy, T.E. (1985) The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain* 23:345-356.

Kersten, P., Kűçűkdeveci, A.A., Tennant, A. (2012) The use of visual analogue (vas) scales in rehabilitation outcomes. *Journal of Rehabilitation Medicine* 44:609-610.

Khan, M. (2011) Different Approaches to White Box Testing Technique for Finding Errors. *International Journal of Software Engineering and Its Applications*. 5(3).

Kirshner, B., Guyatt, G. (1985) A methodological framework for assessing health indices. *Journal of Chronic Disease* 38:27-36.

Kitson, A., Harvey, G., McCormack, B. (1998) Enabling the implementation of evidence based practice: a conceptual framework. *Qual Health Care.* 7(3):149-58.

Kline, C. A. (1965). Osteopathic manipulative therapy, antibiotics, and supportive therapy in respiratory infections in children: comparative study. *J Am Osteopath Assoc.* 65(3): 278-281.

Kling, R., Scacchi, W. (1982) The web of computing: Computer technology as social organisation, In: MC Yovitz, editor *Advances in Computers*, Vol 21 (New York: Academic Press.

Knowles, S., Lam, L.T., McInnes, E., Elliott, D., Hardy, J., Middleton, S. (2015) Knowledge, attitudes, beliefs and behaviour intentions for three bowel management practices in intensive care: effects of a targeted protocol implementation for nursing and medical staff. *BMC Nursing* 14:6

Knox, S.A., King, M.T. (2009) Validation and calibration of the SF-36health transition question against an external criterion of clinical change in health status. *Quality of Life Research*18(5):637-45.

Knutson, G. A. (2003). Vectored upper cervical manipulation for chronic sleep bruxism, headache, and cervical spine pain in a child*. J Manipulative Physiol Ther*. 26(6): E16.

Koch, R. (1882) Die Aetiologie der Tuberculose. *Berliner Klinische Wochenschrift* 19 : 221-230.

Koch, R. (1884) Sechster Bericht der Leiters der deutschen wissenschaftlichen Commission zur Erforschung der Cholera. *Deutsche medizinische Wochenscrift* 10 (12) : 191-192.

Kok, G., de Vries, H., Mudde, A. N., & Strecher, V. J. (1991). Planned health education and the role of self-efficacy: Dutch research. *Health Education Research*, *6* (2), 231-238.

Kompoliti, K., Fan, W., Leurgans, S. (2009). Complementary and alternative medicine use in Gilles de la Tourette syndrome. *Mov Disord*. 24(13): 2015-2019.

Kotronoulas, G., Kearney, N., Maguire, R. (2014) What is the value of the routine use of patient reported outcome measures toward improvement of patient outcomes, processes of care, and health services outcomes in cancer care? *Journal of Clinical Oncology* 32(14):1480-1501.

Kotz, D. (2011) *A threat taxonomy for mHealth privacy*. In: Third International Conference on Communication Systems and Networks (COMSNETS).

KPMG (Klynveld Peat Marwick Goerdeler). (2011) *Report A: How do osteopaths practise?* March, 2011. Retrieved 20th June,2015 from http://www.osteopathy.org.uk/news-and-resources/document-library/continuing-fitness-to-practise/kpmg-report-a-how-do-osteopaths-practise-ozone/.

Kravitz, R., Callahan, E., Azari, R. (1997) Assessing patients' expectations in ambulatory medical practice. *Journal of General Internal Medicine* 12: 67-72.

Kremer, E., Atkinson, J.H., Ignelzi, R.J. (1981) Measurement of pain: patient preference does not confound pain measurement. *Pain* 10:241-248.

Krumholtz, H.M. (2009) Outcomes research: Myths and Realities. *Circulation* 2:1-3.
Kusurkar, R.A., Croiset, G., Mann, K.V., Custers, E., Ten Cate, O. (2012) Have motivation theories guided the development and reform of medical education curricula? A review of the literature. *Academic Medicine* 87(6):735-43.

Laing, R. (1971). *The Politics of the Family and Other Essays*. London; Tavistock Publications.

Landsberger, H.A. (1958). *Hawthorne Revisited.* Ithaca, NY. Cornell University. Social Forces.

Langworthy, J., A. Breen, Vogel, S. Collier, R. Sutherland, G. (2000). *Manipulation Services for NHS Patients: Precedents and Future Models for Provision.* Bournemouth, Institute for Musculoskeletal Research and Clinical Implementation**:** i-vii,1-37.

Lara-Muñoz, C., De Leon, S.P., Feinstein, A.R., Puente, A., Wells, C.K. (2004) Comparison of three rating scales for measuring subjective phenomena in clinical research. I. Use of experimentally controlled auditory stimuli. *Archives of Medical Research* 35(1):43-8.

Lassere, M.N., van der Heijde, D., Johnson, K.R. (2001) Foundations of the minimal clinically important difference for imaging. *Journal of Rheumatology* 28: 890-891.

Lateef, F. (2011) Patient expectations and the paradigm shift of care in emergency medicine. *J Emerg Trauma Shock*. 4(2):163-7.

Lauridsen, H.H., Hartvigsen, J., Korsholm, L., Grunnet-Nilsson, N., Manniche, C. (2007) Choice of external criteria in back pain research: does it matter? Recommendations based on analysis of responsiveness. *Pain* 131;(1-2):112-20.

Lazarus,R.S., Folkman,S. (1984) Stress, appraisal and coping. New York: Springer-Verlag.

Leach, C.M.J., Mandy, A., Hankins, M., Ives, R., Cross, V., Cage, M., Lucas, K. (2011) Communicating risks of treatment and informed consent in osteopathic practice: A literature review and pilot focus groups. Retrieved 12th February, 2015 from http://www.ncor.org.uk/wp-content/uploads/2012/10/communicating-risk.pdf.

Leach, C.M.J., Fiske, A., Mullinger, B., Ives, R., Mandy, A. (2011) Complaints and claims against osteopaths: a baseline study of the frequency of complaints 2004–2008 and a qualitative exploration of patients' complaints. Retrieved 12th February, 2015 from http://www.osteopathy.org.uk/uploads/complaints_and_claims_against_osteopaths_2004-2008_public.pdf.

Leach, C.M.J., Mandy, A., Hankins, M., Bottomley, L.M., Cross, V., Fawkes, C.A., Fiske, A., Moore, A/P. (2013) Patients' expectations of private osteopathic care in the UK: a national survey of patients. *BMC Complementary and Alternative Medicine* 13:122.

Leclaire, R., Blier, F., Fortin, L., Proulx, R. (1997) A cross-sectional study comparing the Oswestry and Roland-Morris Functional Disability scales in two populations of patients with low back pain of different levels of severity. *Spine* (Phila Pa 1976). 1;22(1):68-71.

Lee, S.J., Kavanagh, A. (2007) Electronic and computer-generated patient questionnaires in standard care. *Best Practice and Research Clinical Rheumatology* 7;21(4):637-647.

Leplège, A., Hunt, S. (1997) The problem of quality of life in medicine. *Journal of the American Medical Association* 278:47-50.

Lesho, E.P. (1999) An overview of osteopathic medicine. *Archives of Family Medicine* 8,477-84.

Lewin, S., Glenton, C., Oxman, A.D. (2009) Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. *British Medical Journal* 339:b3496.

Lewis, A.P., Bolden, K.J. (1989) General practitioners and their learning styles. *Journal of the Royal College of General Practitioners* 39, 187-9.

Li, M., Harris, I., Lu, Z.K. (2015) Differences in proxy-reported and patient-reported outcomes: assessing health and functional status among medicare beneficiaries. BMC Med Res Methodol. 15:62.

Licciardone, J.C., Stoll, S.T., Fulda, K.G. Osteopathic manipulative treatment for chronic low back pain: a randomized controlled trial. *Spine* 28:1355-1362.

Licciardone, J.C., Brimhall, A.K., King, L.N. (2005) Osteopathic manipulative treatment for low back pain: a systematic review and meta-analysis of randomized controlled trials. *BMC Musculoskeletal Disorders* 6:43.

Licciardone, J.C., King, H.H., Hensel, K.L. (2008) OSTEOPAThic Health outcomes in chronic low back pain: The OSTEOPATHIC Trial. *Osteopath Medicine in Primary Care* 2:5.

Licciardone, J.C., Buchanan, S., Hensel, K.L. (2009) Osteopathic manipulative treatment of back pain and related symptoms during pregnancy: a randomized controlled trial. *American Journal of Obstetrics and Gynecology* 202(1):43.e1-8.

Licciardone, J.C., Kearns, C.M., Crow, W.T. (2014a) Changes in biomechanical dysfunction and low back pain reduction with osteopathic manual treatment: Results from the OSTEOPATHIC Trial. *Manual Therapy* 19(6):541-8.

Licciardone, J.C., Kearns, C.M., King, H.H. (2014b) Somatic dysfunction and use of osteopathic manual treatment techniques during ambulatory medical care visits: a

CONCORD-PBRN study. *Journal of the American Osteopathic Association* 114(5):344-54.

Licciardone, J.C., Gatchel, R.J., Aryal, S. (2016). Recovery From Chronic Low Back Pain After Osteopathic Manipulative Treatment: A Randomized Controlled Trial. *Journal of the American Osteopathic Association* 116(3):144-55.

Lin, C.W., McAuley, J.H., Macedo, L., Barnett, D.C., Smeets, R.J., Verbunt, J.A. (2011). Relationship between physical activity and disability in low back pain: a systematic review and meta-analysis. *Pain* 152(3):607-13.

Lin, Y-C., Su, Y-C., Chen, Y-P. (2014) Developing Mobile BIM/2D Barcode-Based Automated Facility Management System. The Scientific World Journal http://dx.doi.org/10.1155/2014/374735

Lincoln, Y.S., Guba, E.G. (1985). Naturalistic inquiry, Sage: Beverly Hills.

Lind, J. (1772) *A treatise on the scurvy. 3rd edn.* London: S Crowder.

Lingard L, Albert M, Levinson W. Grounded theory, mixed methods, and action research. *BMJ* 2008;337:a567

Lisi, A.J., Holmes, E.J., Ammendolia, C. (2005) High-velocity low-amplitude spinal manipulation for symptomatic lumbar disk disease: a systematic review of the literature. *Journal of Manipulative and Physiological Therapeutics* 28(6): 429-442.

Litcher-Kelly, L., Martino, S.A., Broderick, J.E., Stone, A.A. (2007). A systematic review of measures used to assess chronic musculoskeletal pain in clinical and randomized controlled clinical trials. *Journal of Pain* 8(12):906-913.

Little, P., Dorward, M., Dorward, G. (2004) Importance of patient pressure and perceived medical need for investigation, referral, and prescription in primary care: Nested observational study. *British Medical Journal*. 328: 444-6.

Little, D.G., MacDonald, D. (1994) The use of the percentage change in Oswestry Disability Index score as an outcome measure in lumbar spinal surgery. *Spine* (Phila Pa 1976). 19(19):2139-43.

Lohr, K.N. (1992). Applications of health status assessment measures in clinical practice. Overview of the third conference on advances in health status assessment. *Medical Care*, 30:MS1-MS4.

Lohr, K.N., Aaronson, N.K., Alonso, J., Burnam, M.A., Patrick, D.L., Perrin, E.B., Roberts, J.S. (1996) Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical Therapeutics* 18:979–992.

Longley, S., Singleton, A. (2008) Social deprivation and digital exclusion in England. *University College London Working Papers Series* 145.

Longo, U.G., Loppini, M., Denaro, L., Maffulli, N., Denaro, V. (2010) Rating scales for low back pain. *British Medical Bulletin* 94:81-144.

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. *Family Practice*, 27:609-614.

Luckett, T., Butow, P.N., King, M.T (2009). Improving patient outcomes through the routine use of patient-reported data in cancer clinics: future directions. *Psycho-Oncology*, 18: 1129–1138.

Lydick, E., Epstein, R.S. (1993) Interpretation of quality of life changes. *Quality of Life Research* 2(3):221-6.

Lydick, E. (2000) Approaches to the interpretation of quality-of-life scales. *Medical Care* 38(9 Suppl):II180-3.

Macedo, L.G., Maher, C.G., Latimer, J., Hancock, M.J., Machado, L.A., McAuley, J.H. (2011) Responsiveness of the 24-, 18- and 11-item versions of the Roland Morris Disability Questionnaire. *European Spine Journal* 20(3):458-63.

MacCarthy, D., Kallstrom, L., Kadlec, H., Hollander, M. (2012) Improving primary care in British Columbia, Canada: evaluation of a peer-to-peer continuing education program for family physicians. *BMC Medical Education* 12:110.

Macfarlane, J., Homes, W., Macfarlane, R. (1997) Influence of patients' expectations on antibiotic management of acute lower respiratory tract illness in general practice: Questionnaire study. *British Medical Journal* 315(7117):1211-4.

Macfarlane, J., Homes, W. (2006) Influence of patients' expectations on antibiotic management of acute lower respiratory tract illness in general practice: Questionnaire study. *British Medical Journal* 315: 1211-1214.

Magni, G., Marchetti, M., Moreschi, C. (1993) Chronic musculoskeletal pain and depressive symptoms in the national health and nutrition examination: Epidemiological follow up study. *Pain* 53: 163-168.

Magruder-Habib, K., Zung, W.W.K., Fuessner, J.R. (1990). Improving physicians' recognisition and treatment of depression in general medical care. Results from a randomised controlled trial. *Medical Care*, 28:239-50.

Mahlknecht, A., Abuzahra, M.E., Piccoliori, G., Enthaler, N., Engl, A., Sönnichsen, A. (2016). Improving quality of care in general practices by self-audit, benchmarking, and quality circles. *Wien Klin Wochenschr,* 128(19-20):706-718.

Maibach, E., Murphy, D.A. (1995) Self-efficacy in health promotion research and practice: conceptualization and measurement. *Health Educ. Res***.** 10 (1):37-50.

Main, C.J. (1983) The Modified Somatic Perception Questionnaire (MSPQ). *Journal of Psychosomatic Research* 27(6):503-14.

Main, C.J., Wood, P.L.R., Hollis, S., Spanswick, C.C., Waddell, G. (1992) The distress and risk assessment method. A simple patient classification to identify distress and evaluate the risk of poor outcome. *Spine* 17:42-52.

Malhotra, K., Buraimoh, O., Thornton, J., Cullen, N., Singh, D., Goldberg, A.J. (2016) Electronic capture of patient-reported and clinician-reported outcome measures in an elective orthopaedic setting: a retrospective cohort analysis. British Medical Journal Open 6(6):e011975.

Malinowski, B. (1932). Crime and custom in savage society. London: Paul, Trench, Trubner.

Malmivaara, A., Hakkinen, U., Heinrichs, M., Koskenniemi, L., Kuosma, E., Lappi, S., Paloheimo, R., Sero, C., Vaaranen, V., Hernberg, S. (1995). The treatment of acute low back pain – bed rest, exercises, or ordinary activity? *New England Journal of Medicine*, 332:351-5.

Mancuso, C.A., Charlson, M.E. (1995) Does recollection error threaten the validity of cross-sectional studies of effectiveness? *Medical Care* 33:AS77-AS88.

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

Mannion, A.F., Junge, A., Grob, D. (2006) Development of a German version of the Oswestry Disability Index. Part 2: sensitivity to change after spinal surgery. *European Spine Journal* 15(1):66-73.

Mansouri, N., Gharaee, B., Shariat, S.V., Bolhari, J., Nooraie, R.Y., Rahimi-Movaghar, A., Alirezaie, N. (2009) The change in attitude and knowledge of health care personnel and general population following trainings provided during integration of mental health in Primary Health Care in Iran: a systematic review. *International Journal of Mental Health Systems* 3:15.

MAPI Research Institute www.mapi-research-inst.com

Mars, T., Ellard, D., Carnes, D., Homer, K., Underwood, M., Taylor, S.J. (2013) Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity. *British Medical Journal Open* 3(11):e003555.

Marshall, S., Haywood, K., Fitzpatrick, R. (2006) Impact of patient-reported outcome measures on routine practice: a structured review. *Journal of Evaluation in Clinical Practice* 12:559-68.

Martel, J., Dugas, C., Lafond, D., Descarreaux, M. (2009) Validation de la version Française du Questionnaire de Bournemouth. *Journal of the Canadian Chiropractic Association* 53(2):111-120.

Marton, F. (1986). Phenomenography - A research approach investigating different understandings of reality. *Journal of Thought* 21(2), 28-49.

Mason, M. Sample Size and Saturation in PhD Studies Using Qualitative Interviews [63 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 2010; 11(3), Art. 8. Retrieved 22nd January, 2014 from http://nbn-resolving.de/urn:nbn:de:0114-fqs100387. at http://www.qualitative-research.net/index.php/fqs/article/view/1428/3027.

Maughan, E.F., Lewis, J.S. (2010) Outcome measures in chronic low back pain. *European Spine Journal* 19(9):1484-1494.

Mausner, J.S., Kramer, S. (1985) In Mausner& Bahn—*Epidemiology: An Introductory Text*. Philadelphia: W B Saunders.

Mauss, M. (1967). The Gift. New York: WW Norton.

Maxwell, J.A. (1996) *Qualitative Research Design: An Interactive Approach* Thousand Oaks, CA: Sage Publications.

May, E.L. (2003) The case for bar coding: better information, better care--and better business. *Healthc Exec*. 18(5):8-13.

McAuley, C., Westby, M.D., Hoens, A., Troughton, D., Field, R., Duggan, M., Reid, W.D. (2014) A survey of physiotherapists' experience using outcome measures in total hip and knee arthroplasty. *Physiotherapy Canada* 66(3):274-85.

McCaffrey, M., Beebe, A. (1989) Giving narcotics for pain. *Nursing* 19:161-165.

McCambridge, J., Witton, J., Elbourne, D.R. (2014) Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol.* 67(3):267-77.

McCoyd, .JL., Kerson, T.S. (2006) Conducting intensive interviews using email: A serendipitous comparative opportunity. *Qualitative Social Work* 5:389–406.

McDowell, I. (1987). *Measuring Health: A Guide to Rating Scales and Questionnaires*, Volume 1. Oxford: Oxford University Press.

McDowell, I. (2006) *Measuring Health.* 3rd ed. Oxford: Oxford University Press.

McGee, J. (1999) *Writing and Designing Print Materials for Beneficiaries: A Guide for State Medicaid Agencies*. Health Care Financing Administration, Center for Medicaid and State Operations, Baltimore, MD.

McGrath, C., Comfort, M.B., Lo, E.C., Luo, Y. (2003) Patient-centred outcome measures in oral surgery: validity and sensitivity. *Br J Oral Maxillofac Surg*. 41(1):43-7.

McGraw, K.O., Wong, S.P. (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1(1): 30-46

McGrew, J., Griss, M. (2005) Concurrent and predictive validity of two scales to assess the fidelity of implementation of supported employment. *Psychiatr Rehab J.* 29: 41-47.

McHorney, C.A. (1999). Health status assessment methods for adults: past accomplishments and future challenges. *Annual Review of Public Health*, 20:309-335.

McIlwraith, B. (2003) A survey of 1200 osteopathic patients in the United Kingdom. *Journal of Osteopathic Medicine* 6(1): 7-12.

McInerney, S.J. (2015) Introducing the Biopsychosocial Model for good medicine and good doctors. *British Medical Journal* 324:1533.

McIver, S. (1991a) *Obtaining the views of outpatients*. Kings Fund, London

McIver, S. (1991b) *An introduction to obtaining the views of users in health services*. London: Kings Fund.

McNee, P., Shambrook, J., Harris, E.C., Kim, M., Sampson, M., Palmer, K.T., Coggon, D. (2011). Predictors of long term pain and disability in patients with low back pain investigated by magnetic resonance imaging: a longitudinal study. *BMC Musculoskeletal Disorders*, 12:234.

Meadows, K.A., Rogers, D., Greene, T. (1998). Attitudes to the use of health outcome questionnaires in the routine care of patients with diabetes: a survey of general practitioners and practice nurses. *The British Journal of General Practice*, 48: 1555-1559.

Means, R., Smith, R. (1998) *Community Care: Policy and Practice*, Second Edition, Basingstoke: MacMillan.

Med Confidential http://medconfidential.org/whats-the-story/

Medical outcomes Trust http://www.outcomes-trust.org/

Medical Research Council ethics series (2012). *Good Research Practice: principles and guidance*. Retrieved 20th January, 2014 from http://www.mrc.ac.uk/publications/browse/good-research-practice-principles-and-guidelines/.

Meinert, R., Yuen, F.K.O. (2012) *Controversies and Disputes in Disability and Rehabilitation*. Oxford: Routledge.

Meissner, W., Ullrich, K., Zwacka, S. (2006). Benchmarking as a tool of continuous quality improvement in postoperative pain management. *European Journal of Anaesthesia*, 23(2):142-8.

Meissner, W. (2008). Continuous Quality Improvement in Postoperative Pain Management – benchmarking post-operative pain. *Acute Pain*, 8:43-44.

Melzack, R. (1975) The McGill Pain Questionnaire: major properties and scoring methods. *Pain.* 1(3):277-99.

Menezes da Costa, L., Maher, C.G., McAuley, J.H., Costa, L.O. (2009). Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing. *Journal of Clinical Epidemiology* 62:934-943.

Menke, J.M. (2014). Do manual therapies help low back pain? A comparative effectiveness meta-analysis. *Spine* 39(7):E463-72.

Merton, R.K., Fiske, M., Kendall, P.L. (1956) *The Focused Interview*. Glencoe, Illinois: Free Press.

MHRA (2014) Yellow card scheme. https://yellowcard.mhra.gov.uk/the-yellow-card-scheme/

Microsoft. (2005) https://www.microsoft.com/presspass/newsroom/winxp/WinVistaCTPFS.mspx

Mid Staffordshire Enquiry (2014). Retrieved on 14-06-2015 from https://www.gov.uk/government/collections/mid-staffordshire-nhs-foundation-trust-news-and-publications

Mid-Staffordshire NHS Foundation Trust. *Public Enquiry* . Retrieved 13th May, 2014 from http://www.midstaffspublicinquiry.com/.

Mihalic, S. (2002) The importance of implementation fidelity. *Blueprints Violence Prevention Initiative.* Boulder, Colorado.

Mihalic, S. (2004) The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth.* 4: 83-86. and 99–105.

Miller, J.N., Colditz, G.A., Mosteller, F. (1989) How study design affects outcomes in comparisons of therapy. II. Surgical. *Statistics in Medicine* 8: 455–466.

Miller, J. E., Newell, D., Bolton, J.E. (2009). Contribution of chiropractic therapy to resolving suboptimal breastfeeding: a case series of 114 infants. *J Manipulative Physiol Ther*. 32(8): 670-674

Miller, J. E., Phillips, H.L. (2009). Long-term effects of infant colic: a survey comparison of chiropractic treatment and nontreatment groups. *J Manipulative Physiol Ther*. 32(8): 635-638.

Moffett, J.A., Underwood, M.R., Gardiner, E.D. (2009). Socioeconomic status predicts functional disability in patients participating in a back pain trial. *Disability and Rehabilitation*, 31(10):783-790.

Mogren, I.M. (2006). BMI, pain and hyper-mobility are determinants of long-term outcome for women with low back pain and pelvic pain during pregnancy. *European Spine Journal*, 15(7):1093-1102.

Moher, D., Liberati, A., Tetzlaff, J., and The PRISMA Group. (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7): e1000097.

Mokkink, L.B., Terwee, C.B., Knol, D.L., Stratford, P.W., Alonso, J., Patrick, D.L., Bouter, L.M., de Vet, H.C.W. (2006) Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Research Methodology* 6:2-9.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., De Vet, H.C.W. (2010a) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research* 19(4):539-49.

Mokkink, L.B., Terwee, C.B., Knol, D.L., Stratford, P.W., Alonso, J., Patrick, D.L., Bouter, L.M., De Vet, H.C.W. (2010b) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Medical Research Methodology* 10:22.

Molm, L.D. (1994) Is punishment effective: coercive theories in social exchange. Social Psychology Quarterly 57;75-94.

Molm, L. D. (1997). Coercive power in social exchange. Cambridge, UK: Cambridge University Press.

Molm, L. D. (2000). *Theories of social exchange and exchange networks*. In G. Ritzer & B. Smart (Eds.), Handbook of social theory: 260-272. Thousand Oaks, CA: Sage

Molm, L.D. (2003). Theoretical Comparisons of Forms of Exchange. *Sociological Theory* 21(1):1-17.

Monaco, A., Cozzolino, V., Cattaneo, R., Cutilli, T., Spadaro, A.. (2008). Osteopathic manipulative treatment (OMT) effects on mandibular kinetics: kinesiographic study. *Eur J Paediatr Dent*. 9(1): 37-42.

Monopolies and Mergers Commission (1989) Services of professionally regulated osteopaths. Cm583.

Moore, A., Jull, G. (2006) The systematic review of systematic reviews has arrived! *Manual Therapy* 11:91–92.

Moore, A.P., Bryant, E.C., Olivier, G.W. (2012) Development and use of standardised data collection tools to support and inform musculoskeletal practice. *Manual Therapy*17(6):489-96.

Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonnell, C., Hardeman, W., Moore, L., O'Cathain, A.O., Tinati, T., Wight, D., Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *British Medical Journal* 350:1258

Moran, R.W., Gibbons, R. (20010 Intraexaminer and Interexaminer Reliability for Palpation of the Cranial Rhythmic Impulse at the Head and Sacrum. *Journal of Manipulative and Physiological Therapeutics* 24:183-90**.**

Moran,P., Kelesidi,K., Guglani,S., Davidson,S., Ford,T. (2012). What do parents and carers think about routine outcome measures and their use? A focus group study of CAMHS attenders. *Clinical Child Psychology and Psychiatry,*17(1):65-79.

Morris, J., Perez, D., McNoe, B. (1998). The use of quality of life data in clinical practice. *Quality of Life Research*, 7: 85-91.

Morse, J.M. (1994). Designing funded qualitative research. In Denzin NK and Lincoln YS (Eds.) *Handbook of qualitative research* (2nd ed). Thousand Oaks, CA: Sage.

M-PROM Briefing. Arthritis Research UK (2013). Retrieved 1st May, 2014 from http://www.arthritisresearchuk.org/policy-and-public-affairs/policy-priorities-and-projects/musculosketal-health-services/patient-reported-outcome-measures.aspx.

M-PROM development Retrieved 1st May, 2014 from http://public.ukcrn.org.uk/search/StudyDetail.aspx?StudyID=10511.

Mullen, P.M. (1999) Public involvement in healthcare priority setting: an overview of methods for eliciting values. *Health Expectations* 2: 222-34.

Müller, U., Duetz, M.S., Roeder, C., Greenough, C.G. (2004) Condition-specific outcome measures for low back pain. Part I: validation. *Eur Spine J*. ;13(4):301-13.
Mulrow, C.D. (1987) The medical review article: state of the science. *Ann Int Med* 106:485-8

Murphy, D.R., Hurwitz, E.L., McGovern, E.E. (2009) Outcome of pregnancy-related lumbopelvic pain treated according to a diagnosis-based decision rule: a prospective observational cohort study. *J Manipulative Physiol Ther.* 32(8):616-24.

Murphy, D.R., Hurwitz, E.L. (2011) The usefulness of clinical measures of psychologic factors in patients with spinal pain. *J Manipulative Physiol Ther.* 34(9):609-13.

Murphy, D.R., Lopez, M. (2013). Neck and back pain specific outcome assessment questionnaires in the Spanish language: a systematic literature review. *Spine Journal* 13(11):1667-74.

Murphy, M., Hollinghurst, S., Salisbury, C. (2014) Can the outcome of primary care be measured by a Patient Reported Outcome Measure? *British Journal of General Practice* 64(629)647-8.

Murphy, M., Hollinghurst, S., Turner, K., Salisbury, C. (2015) Patient and practitioners' views on the most important outcomes arising from primary care consultations: a qualitative study. *BMC Family Practice* 16**:**108.

Murphy, M., Hollinghurst, S., Salisbury, C. (2016a) Agreeing the content of a patient-reported outcome measure for primary care: a Delphi consensus study. *Health Expectations* 20**:**335-348.

Murphy, M., Hollinghurst, S., Salisbury, C. (2016b) Development of the Primary Care Outcomes Questionnaire (PCOQ). In Proceedings of Patient Reported Outcome Measure's (PROMs) Conference Sheffield 2016: advances in patient reported outcomes research. *Health and Quality of Life Outcomes* 14 (Suppl 1)**:**137-143.

Murray, E. (2007) Using the Internet for research: results at a keystroke. *British Journal of General Practice* 57:939-40.

My Clinical Outcomes. www.myclinicaloutcomes.co.uk.

Myers, S.S., Phillips, R.S., Davis, R.B., Cherkin, D.C., Legedza, A., Kaptchuk, T.J., Hrbek, A., Buring, J.E., Post, D., Connelly, M.T. (2008) Patient expectations as predictors of outcome in patients with acute low back pain. *Journal of General Internal Medicine* 23 (2): 148-153.

Nabukera, S. K., Romitti, P.A., Campbell, K.A., Meaney, F.J., Caspers, K.M., Mathews, K.D., Sherlock, S.M., Puzhankara, S., Cunniff, C., Druschel, C.M., Pandya, S., Matthews, D.J., Ciafaloni, E.; MD STARnet. (2012). Use of complementary and alternative medicine by males with Duchenne or Becker muscular dystrophy. *J Child Neurol*. 27(6): 734-740.

Nachemson, A., Bigos, S.J. (1984) The low back. In: Creuss J, Rennie WJR (eds) *Adult orthopaedics*. New York: Churchill-Livingstone.

National Audit Office. Report on NHS IT programme for delivery of detailed care records system. Retrieved 1st May, 2014 from http://www.nao.org.uk/report/the-national-programme-for-it-in-the-nhs-an-update-on-the-delivery-of-detailed-care-records-systems/.

National Data Guardian for Health and Care. (2016) Review of Data Security, Consent and Opt-outs.

National Institute of Health and Care Excellence (2014) *Updating of CG88: Low back pain management of persistent non-specific low back pain.* Retrieved 1st May, 2014 from http://publications.nice.org.uk/low-back-pain-cg88/updating-the-guideline.

National Institute for Health and Care Excellence (NICE) (2016). Low back pain and sciatica. https://www.nice.org.uk/guidance/indevelopment/gid-cgwave0681

National PROMs programme in the NHS (2014) Retrieved 1st May, 2014 from http://www.hscic.gov.uk/proms.

Newman, A.N.L., Stratford, P.W., Letts, L., Spadoni, G. (2013) A systematic review of Head-to-Head comparison studies of the Roland-Morris and Oswestry Measures' abilities to assess change. *Physiotherapy Canada* 65(2):160-166.

Nelson, E.C., Conger, B., Douglass, R., Gephart, D., Kirk, J., Page, R. (1983) Functional health status levels of primary care patients. *Journal of the American Medical Association* 249:3331-8.

Nelson, E., Wasson, J., Kirk, J. (1987) Assessment of function in routine clinical practice: description of the COOP-Chart method and preliminary findings. *Journal of Chronic Disease* 40 (Suppl. 1):55S–63S.

Nelson, E.C., Eftimovska, E., Lind, C., Hager, A., Wasson, .JH., Lindblad, S. (2015) Patient reported outcome measures in practice. *British Medical Journal* 350:7818.

Neroev, V. V., Chuvilina, M.V., Tarutta, E.P., Ivanov, A.N. (2006). Reflex therapy, massage, and manual therapy in the treatment of progressive myopia in children and adolescents. *Vestn Oftalmol*. 122(4): 20-24.

Nevo, I., Slonim-Nevo, V. (2011) The myth of evidence-based practice: towards evidence-informed practice. *British Journal of Social Work* 1-22.

Newell, D., Bolton, J.E. (2010) Responsiveness of the Bournemouth questionnaire in determining minimal clinically important change in subgroups of low back pain patients. *Spine* (Phila Pa 1976). 1;35(19):1801-6.

Newman, A.N., Stratford, P.W., Letts, L., Spadoni, G. (2013). A Systematic Review of Head-to-Head Comparison Studies of the Roland-Morris and Oswestry Measures' Abilities to Assess Change. *Physiotherapy Canada* 65(2):160-6.

Newton, J.N., Briggs, A.D.M., Murray, C.J.L., Dicker, D., Foreman, K.J., Wang, H., Naghavi, M., Farforouzan, M.H., Ohno, S.L., Barber, R.M., Vos, T., Stanaway, J.D., Schmidt, J.C., Hughes, A.J., Fay, D.F.J., Ecob, R., Gresser, C., McKee, M., Rutter, H., Abubakar, I., Ali, R., Anderson, H.R., Banerjee, A., Bennett, D.A., Bernabé, E., Bhui K.S., Biryukov, S.M., Bourne, R.R., Brayne, C.E.G., Bruce, N.G., Brugha, T.S., Burch, M., Capewell, S., Casey, D., Chowdhury, R., Coates, M.M., Cooper, C., Critchley, J.A., Dargan, P.I., Dherani, M.K., Elliott, P., Ezzati, M., Fenton, K.A., Fraser, M.A., Furst, T., Greaves, M.A., Gunnell, D.J., Hannigan, B.M., Hay, R.J., Hay, S.I., Hemingway, H., Larson, H.J., Davis, A.C.J. Changes in health in England, with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 386(10010):2257-2274.

NHS Confederation (2015) Key statistics on the NHS. Retrieved 14th June, 2016 from http://www.nhsconfed.org/resources/key-statistics-on-the-nhs.

NHS Connecting for Health (2014) Retrieved 1st May, 2014 from http://www.connectingforhealth.nhs.uk/.

NHS Digital (2016) *Provisional Monthly Patient Reported Outcome Measures (PROMs) in England – April 2015 to February 2016*. Retrieved 21st July, 2016 from http://digital.nhs.uk/catalogue/PUB20905.

NHS England Care Data. http://www.england.nhs.uk/ourwork/tsd/care-data/

NHS England, GP toolkit. http://www.england.nhs.uk/gp-toolkit/

NHS England (2015a) Annual Report and Accounts 2014-2015. Retrieved 21st January, 2016 from https://www.england.nhs.uk/wp-content/uploads/2015/07/nhse-annual-report-2014-15.pdf

NHS England (2015b) Health budget 2015-2016. Retrieved 21st January, 2016 from https://www.gov.uk/government/publications/spending-review-and-autumn-statement-2015-documents/spending-review-and-autumn-statement-2015.

NHS Research and Development Forum (2014). http://www.rdforum.nhs.uk/content/resources/#Deciding

Nielsen, J. (1994) Usability Laboratories. *Behavior and Information Technology*, 13(1).

Nightingale, F. (1858) *Notes on matters affecting the health, efficiency, and hospital administration of the British Army [electronic resource]: founded chiefly on the experience of the late war*. London : Harrison. Retrieved 2nd June, 2015 from https://archive.org/stream/b20387118#page/n15/mode/2up.

Njie, L. (2016) Mobile health and fitness apps: what are the privacy risks? Retrieved 6th January, 2016 from https://www.privacyrights.org/mobile-health-and-fitness-apps-what-are-the-privacy-risks.

Nordeman, L., Thorselius, L., Gunnarsson, R., Mannerkorpi, K. (2017). Predictors for future activity limitation in women with chronic low back pain consulting primary care: a 2-year prospective longitudinal cohort study. *BMJ Open*, 7:e013974.

Norman, G.R., Brooks, L.R., Cunningham, J.P. (1996) Expert-novice differences in the use of history and visual information from patients. *Acad Med* 71(10):62-64.
Norman, G.R., Stratford, P., Regehr, G. (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *Journal of Clinical Epidemiology* 50(8):869-879.

Norman, G.R., Sridhar, F.G., Guyatt, G.H., Walter, S.D. (2001) Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Medical Care* 39: 1039-1047.

Novick, G. (2008) Is There a Bias Against Telephone Interviews In Qualitative Research? *Research Nursing and Health* 31(4): 391–398.

Nulty, D.N. (2008) The adequacy of response rates to online and paper surveys: what can be done? *Assessment and Evaluation in Higher Education* 33(3):301-314.

Nutley, S., Homel, P. (2006) Delivering evidence-based policy and practice: Lessons from the implementation of the UK Crime Reduction Programme. *Evidence & Policy*. 2: 5-26.

Oakley, A. (2005) *Design and analysis of social intervention studies in health research.* In: Handbook of Health Research Methods Maidenhead: Open University Press.

Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J., (2006) RIPPLE Study Team. Process evaluation in randomised controlled trials of complex interventions. *British Medical Journal* 332:413-416.

Ofcom (2009) Access and Inclusion - Summary of Ofcom research on internet access, use and attitudes. Retrieved 1st May, 2014 from http://stakeholders.ofcom.org.uk/binaries/consultations/access/annexes/accessandinclusionsummary.pdf

Ofcom research report (2012) Retrieved 1st May, 2014 from http://stakeholders.ofcom.org.uk/binaries/research/consumer-experience/tce-12/Consumer_Experience_Researc1.pdf .

Ofcom (2015) Communications Market report. Retrieved 1st October, 2015 from http://stakeholders.ofcom.org.uk/market-data-research/market-data/communications-market-reports/cmr15/internet-web/

Office for National Statistics (2014) *Full Report: Sickness Absence in the Labour Market.* Retrieved 2nd June, 2015 from http://www.ons.gov.uk/ons/dcp171776_353899.pdf.

Office for National Statistics (ONS) (2008). *All items Retail Price Index (RPI) percentage change over 12 months: Table RP04. 2008.* Retrieved 1st May, 2014 from http://www.ons.gov.uk/ons/rel/cpi/users-and-uses-of-consumer-price-inflation-statistics/2013/index.html.

Office for National Statistics (ONS) (2011) *Census population*. Retrieved 13th May, 2014 from http://www.ons.gov.uk/ons/rel/census/2011-census/population-and-household-estimates-for-england-and-wales/rft-m01.xls.

Office for National Statistics (ONS) (2012) *Cultural identity*. Retrieved 12th May, 2014 from http://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11.

Office for National Statistics (2015). UK Labour Market Statistical Bulletin, July 2015, ONS Table 2(2). Retrieved 1st August, 2015 from http://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/uklabourmarket/2015-07-15.

Office of Fair Trading (1986) Retrieved 14th January, 2014 from https://www.gov.uk/government/organisations/office-of-fair-trading.

Office of the Information Commissioner (2014) Guide to Data Protection. Retrieved 13th June, 2014 from http://ico.org.uk/Global/~/media/documents/library/Data_Protection/Practical_application/THE_GUIDE_TO_DATA_PROTECTION.ashx.

Ogon, M., Krismer, M., Sollner, W., Kantner-Rumplmair, W., Lampe, A. (1996) Chronic low back pain measurement with visual analogue scales in different settings. *Pain* 64:425-428.

Ogus, A.I. (2004) *Regulation: legal form and economic theory*. Oxford: Hart Publishing.

Olafsdottir, E., Forshei, S., Fluge, G., Markestad, T. (2001). Randomised controlled trial of infantile colic treated with chiropractic spinal manipulation. *Arch Dis Child*. 84(2): 138-141

Oliphant, D. (2004) Safety of spinal manipulation in the treatment of lumbar disk herniations: a systematic review and risk assessment. *Journal of Manipulative and Physiological Therapeutics*. 27:197-210.

Oner, F.C., Jacobs, W.C., Lehr, A.M., Sadiqi, S., Post, M.W., Aarabi, B., Chapman, J.R., Dvorak, M.F., Fehlings, M.G., Kandziora, F., Rajasekaran, S., Vaccaro, A.R. (2016). Toward the Development of a Universal Outcome Instrument for Spine Trauma: A Systematic Review and Content Comparison of Outcome Measures Used in Spine Trauma Research Using the ICF as Reference. *Spine* (Phila Pa 1976). 41(4):358-67.

Ong, C-K . (2004) Use of osteopathic or chiropractic services among people with back pain: a UK population survey. *Health and Social Care in the Community* 12(3):265-273

OPCS report. (1994) *The prevalence of back pain in Great Britain*. London: HMSO. Orthopaedic scores http://www.orthoscores.com/ .

Ostelo, R.W.J.G., de Vet, H.C.W. (2005) Clinically important outcomes in low back pain. Best Practice and Research. *Clinical Rheumatology* 19(4):593-607.

Ostelo, R.W., de Vet, H.C., Knol, D.L. (2004) 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *Journal of Clinical Epidemiology* 57:268–76.

Ostelo, R.W., Deyo, R.A., Stratford, P., Waddell, G., Croft, P., Von Korff, M., Bouter, L.M., de Vet, H.C. (2008) Interpreting change scores for pain and functional status in

low back pain: towards international consensus regarding minimal important change. *Spine* (Phila Pa 1976). 33(1):90-4.

Osteopath 's Act (1993). Retrieved 1st May, 2014 from http://www.opsi.gov.uk/acts/acts1993/Ukpga_19930021_en_1.

Outcome Measures in Rheumatology Clinical Trials (OMERACT). Retrieved 6th June, 2015 from http://www.omeract.org/.

Overton, G.K., Kelly, D., McCalister, P., Jones, J., MacVicar, R. (2009) The practice-based small group learning approach: making evidence-based practice come alive for learners. *Nurse Education Today* 29(6):671-5.

Owens, C. (1963) *Endocrine Interpretation of Chapman 's Reflexes*. Newark, Ohio: American Academy of Osteopathy.

Oxman A.D., Guyatt, G.H. (1988) Guidelines for reading literature reviews. *Can Med Assoc J* 138:697-703.

Oxman, A.D., Thomson, M.A., Davis, D.A., Haynes, R.B. (1995) No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ.* 153(10):1423-31.

Palmer, A. (2012) *An introduction to marketing theory and practice*. Oxford; OUP.

Pannucci, C.J., Wilkins, E.G. (2010) *Identifying and Avoiding Bias in Research. Plastic Reconstructive Surgery* 126(2): 619 –625.

Parker, S.L., Adogwa, O., Paul, A.R., Anderson, W.N., Aaronson, O., Cheng, J.S., McGirt, M.J. (2011). Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. *Journal of Neurosurgery: Spine* 14(5):598-604.

Parsons, S., Harding, G., Breen, A. (2011) The influence of patients' and primary care practitioners beliefs and expectations about chronic musculoskeletal pain on the process of care: a systematic review of qualitative studies. *Clinical Journal of Pain.* 23(1):91-98.

Pascoe, G.C. (1983) Patient satisfaction in primary healthcare: a literature review and analysis. *Evaluation and Program Planning* 6:185 -210.

Pasteur, L. (1858). Nouveaux faits concernant l'histoire de la fermentation alcoolique. *Comptes Rendus Chimie* 47: 1011–1013.

Patient Reported Outcome and Quality of Life Health Measurement Database (PROQOLID). Retrieved 1st May, 2014 from http://www.proqolid.org/ http://www.rehabmeasures.org/default.aspx .

Patrick, D.L., Deyo, R., Atlas, S., Singer, D., Chapin, A., Keller, R. (1995) Assessing health related quality of life in patients with sciatica. *Spine* 20:1899-1908.

Patrick, D.L., Guyatt, G.H., Acquadro, C. (2008) On behalf of the Cochrane Patient Reported Outcomes Methods Cochrane Review Group. Patient Reported Outcomes. Ch. 17. Retrieved 1st June, 2014 from http://hiv.cochrane.org/sites/hiv.cochrane.org/files/uploads/Ch17_PRO.pdf.

Patrick, D.L., Burke, L.B., Gwaltney, C.J. (2011a) Content Validity —Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1—Eliciting Concepts for a New PRO Instrument. *Value in Health* 14:967-977.

Patrick, D.L., Burke, L.B., Gwaltney, C.J. (2011b) Content Validity —Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2—Assessing Respondent Understanding. *Value in Health* 14:978-988.

Peabody, J.W., Pacudo, D.Ouenes, O., Shimkhada, R., DeMaria, L., Burgon, T.B. (2016). Improving clinical practice using a novel engagement approach: measurement, benchmarking, and feedback, a longitudinal study. *Journal of Clinical Medicine Research*, 8(9):633-640.

Pearson, K. (1904) Report on certain enteric fever inoculation statistics. *BMJ* 3:1243–6.

Pengel, L.H.M., Herbert, R.D., Maher, C.G., Refschauge, K.M. (2003). Acute low back pain: a systematic review of its prognosis. *British Medical Journal*, 327:323-5.

Penuel, W., Means, B. (2004) Implementation variation and fidelity in an Inquiry Science Program: analysis of GLOBE data reporting patterns. *J Res Sci Teaching*. 41: 294-315.

Perillo, M., Bulbulian, R. (2003) Responsiveness of the Bournemouth and Oswestry questionnaires: a prospective pilot study. *J Manipulative Physiol Ther*. 26(2):77-86.

Perrin, E.G. (1995) SAC instrument review process. *Medical Outcomes Trust Bulletin* 3(4):1.

Peterson, C.K., Mühlemann, D., Humphreys, B.K. (2014). Outcomes of pregnant patients with low back pain undergoing chiropractic treatment: a prospective cohort study with short term, medium term and 1-year follow up. *Chiropractic Manual Therapy*, 22(1):15.

Peute, L.W., Aarts. J., Bakker, P.J., Jaspers, M.W. (2010) Anatomy of a failure: a sociotechnical evaluation of a laboratory physician order entry system implementation. *International Journal of Medical Informatics* 79(4):e58-70.

Philippi, H., Faldum, A., Schleupen, A., Pabst, B., Jung, T., Bergmann, H., Bieber, I., Kaemmerer, C., Dijs, P., Reitter, B. (2006). Infantile postural asymmetry and osteopathic treatment: a randomized therapeutic trial. *Dev Med Child Neurol*. 48(1): 5-9

Pincus, T., Vogel, S., Savage, R., Newman, S. (2000) Patient satisfaction with osteopathic and GP management of low back pain in the same surgery. *Complementary Therapies in Medicine* 8(3):180-186.

Pincus, T., Burton, A.K., Vogel, S. (2002) A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* (Phila Pa 1976) 27(5):E109-20.

Pincus, T., Santon, R., Breen. (2008) Review and Proposal for a Core Set of Factors for Prospective Cohorts in Low Back Pain: A Consensus Statement. *Arthritis & Rheumatism (Arthritis Care & Research)* 59(1):14 –24.

Pinnock, H., Epiphaniou, E., Sheikh, A., Griffiths, C., Eldridge, S., Craig, P., Taylor, S.J. (2015). Developing standards for reporting implementation studies of complex interventions (STaRI): a systematic review and e-Delphi. *Journal of Implementation Science* 10:42.

Pinsker, E., Daniels, T.R., Inrig, T. (2013) The ability of outcome questionnaires to capture patient concerns following ankle reconstruction. *Foot and Ankle International* 34(1):65-74.

Pisoni, C., Giardini, A., Majani, G., Maini, M. (2008) International Classification of Functioning, Disability and Health (ICF) core sets for osteoarthritis. A useful tool in the follow-up of patients after joint arthroplasty. *European Journal of Physical and Rehabilitation Medicine* 44(4):377-85.

Ploug, T., Holm, S. (2015). Routinisation of informed consent in online health care systems. *International Journal of Medical Informatics* 84(4):229-36.

Ponce de Leon, S., Lara-Muñoz, C., Feinstein, A.R., Wells, C.K. (2004) A comparison of three rating scales for measuring subjective phenomena in clinical research. II. Use of experimentally controlled visual stimuli. *Archives of Medical Research* 35(2):157-62.

Pope, C., Mays, N. (2006) *Qualitative Research in Healthcare* (3rd Edition) Oxford; BMJ Books

Pope, C., Mays, N. (2009) Critical reflections on the rise of qualitative research. BMJ, 339:b3425 – b3425.

Porter, J. (1998) *Hearing before the subcommittee of the Committee on Appropriations*. One Hundred Fifth Congress, Second Session, Part 3. Washington DC: Department of Health and Human Services.

Posadzki, P., Ernst, E. (2011a) Spinal manipulations for cervicogenic headaches: a systematic review of randomized clinical trials. *Headache* 51(7):1132-9.

Posadzki, P., Ernst, E. (2011b) Spinal manipulation: an update of a systematic review of systematic reviews. *New Zealand Medical Journal* 124(1340):55-71.

Potter, M., Gordon, S., Hamer, P. (2003) The physiotherapy experience in private practice: The patients' perspective. *Australian Journal of Physiotherapy* 49(3):195-202.

Powell-Smith, A., Goldacre, B. (2016) The TrialsTracker: Automated ongoing monitoring of failure to share clinical trial results by all major companies and research. *F1000Research* 2016, 5:2629.

Power, R., Langhaug, L.F., Nyamurera, T., Wilson, D., Bassett, M.T., Cowan, F.M. (2004) Developing complex interventions for rigorous evaluation--a case study from rural Zimbabwe. *Health Educ Res*. 19(5):570-5.

Preston, R.A., Materson, B.J., Reda, D.J., Williams, D.W. (2000). Placebo-associated blood pressure response and adverse effects in the treatment of hypertension:

observations from a Department of Veterans Affairs Cooperative Study. *Archives of Internal Medicine*. 160(10):1449–1454.

Pringle, M., Tyreman, S. (1993) Study of 500 patients attending an osteopathic practice. *British Journal of General Practice* 43: 15-8.

Purse, F. M. (1961). Clinical evaluation of osteopathic manipulative therapy in measles*. J Am Osteopath Assoc.* 61: 274-276.

Purse, F. M. (1966). Manipulative therapy of upper respiratory infections in children. *J Am Osteopath Assoc*. 65(9): 964-972.

Pyett, P. (2003). Validation of qualitative research in the "real world". *Qualitative Health Research* 13:1170–1179.

Radbruch, L., Sabatowski, R., Loick, G., Jonen-Thielemann, I., Kasper, M., Gondek, B., Lehmann, K.A., Thielemann, I. (2000) Cognitive impairment and its influence on pain and symptom assessment in a palliative care unit: development of Minimum Documentation System. *Palliative Medicine* 14:266-276.

Rae, J., Green, B. (2016) Portraying Reflexivity in Health Services Research. *Qualitative Health Research* Mar 1. pii: 1049732316634046. [Epub ahead of print]

Rageot, E. (1968). Complications and accidents in vertebral manipulation. *Cah Coll Med Hop Paris*. 9(14): 1149-1154.

Rajendran, D., Mullinger, B., Fossum, C. (2009) Monitoring self-reported adverse events: A prospective, pilot study in a UK osteopathic teaching clinic. *International Journal of Osteopathic Medicine* 12(2):49-55.

Randell, B., *Fensom, H., Milne, F.A. (1995)* Obituary: Allen Coombs. *The Independent. Retrieved 21st July, 2014 from* https://en.wikipedia.org/wiki/History_of_computing_hardware.

Rankin, P. (2006) The Bournemouth Questionnaire as an outcome measure in the rehabilitation of a person suffering with mechanical neck and arm pain and concurrent Charcot-Marie-Tooth disease: a case report. *J Can Chiropr Assoc.* 50(3):190-4.

Raske, M. (2005). The disability discrimination model in social work practice. In G. E. May & M. B. Raske (Eds.), *Ending disability discrimination: Strategies for social workers*. Boston: Pearson Education, Allyn & Bacon.

Rasmussen-Barr, E., Campello, M., Arvidsson, I., Nilsson-Wikmar, L., Ang, B.O. (2012). Factors predicting clinical outcome 12 and 36 months after an exercise intervention for recurrent low back pain. *Disability and Rehabilitation*, 34(2):136-144.

Raspe, H., Huppe, A., Matthias, C. (2003) Theories and models of chronicity: on the way to a broader definition of chronic back pain. *Schmerz* 17:359-366.

Raw, R.K., Wilkie, R.M., White, A., Williams, J.H., Mon-Williams, M. (2015) The 'Goldilocks Zone': getting the measure of manual asymmetries. *PLoS One*. 10(5):e0128322.

Rawlins, M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions (2008) *Clinical Medicine* 8(6):579-588.

Rayleigh, The Right Hon Lord. (1884) *Presidential address at the 54th meeting of the British Association of Science, Montreal, August/September 1884.* London: John Murray. 3-23.

Rebollo, P., Valderas, J.M., Ortega, F. (2005). Progress in Spain of the described barriers to the use of perceived health status measures in clinical practice. *Medcina Clínica*, 125(18):703-705

Redelmeier, D.A., Lorig, K. (1993) Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. *Archives of Internal Medicine* 153:137-1342.

Reeves, S., Kuper, A., Hodges, B.D. (2008) Qualitative research methodologies: ethnography. *British Medical Journal* 337:a1020.

Reiber, G.E., Au, D., McDonell, M., Fihn, S.D. (2004) Diabetes quality improvement in Department of Veterans Affairs Ambulatory Care Clinics: a group-randomized clinical trial. *Diabetes Care* 27(suppl 2):B61 –8.

Resnick, B., Neale, M., Rosenheck, R. (2003) Impact of public support payments, intensive psychiatric community care, and program fidelity on employment outcomes for people with severe mental illness. *J Nerv Ment Dis*. 191: 139-144.

Richards, N., Coulter, A. (2007). *Is the NHS becoming more patient-centred? Trends from the national surveys of NHS patients in England 2002 –07*. Oxford: Picker Institute Europe.

Richardson, J. (2004) What patients expect from complementary therapy: A qualitative study. *American Journal of Public Health* 94(6): 1049-53.

Riddle, D.L., Stratford, P.W., Binkley, J.M. (1998) Sensitivity to change of the Roland Morris Back Pain Questionnaire: part 2. *Physical Therapy* 78:1197–207.

Riddle, D.L., Stratford, P.W. (2002). Roland Morris scale reliability. *Physical Therapy*, 82(5):512-514.

Ries, E. (2011) *The Lean Start-up. How constant innovation creates radically successful businesses*. London: Penguin.

Riley, M., Richardson, I (1998) An introduction to Reed-Solomon codes: principles, architecture and implementation. Retrieved 14-02-2016 from https://www.cs.cmu.edu/~guyb/realworld/reedsolomon/reed_solomon_codes.html

Ritchie, J., Lewis, J.(eds). (2012) *Qualitative research practice: A guide for social science students and researchers*. London: Sage Publications.

Roberts, A.J., Drew, A., Bridger, R., Etherington, J., Kitminster, S. (2015). Predicting low back pain outcome following rehabilitation for low back pain. *Journal of Back Musculoskeletal Rehabilitation*, 28(1):119-128.

Robert-Lachaine, X., Mecheri, H., Larue, C., Plamondon, A. (2016) Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. *Medical and Biological Engineering and Computing*. Jul 5. [Epub ahead of print]

Roland, M., Fairbank, J. (2000) The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine* 25(24):3115-24.

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

Roland, M., Morris, R. (1983b) A study of the natural history of low-back pain. Part II: development of guidelines for trials of treatment in primary care. *Spine* (Phila Pa 1976) 8(2):145-50.

Roland Morris Disability Questionnaire. Retrieved 1st May, 2013 from http://www.rmdq.org/.

Rosier, E.M., Iadarola, M.J., Coghill, R.C. (2002) Reproducibility of pain measurement and pain perception. *Pain* 98:205-16.

Ross, M. (1989) Relation of implicit theories to the construction of personal histories. *Psychiatric Review* 96:341-57.

Rossi, P., Freeman, H., Lipsey, M. (1999) *Evaluation: a systematic approach*. Thousand Oaks, California: Sage Publications.

Rossignol, M., Begaud, B., Avounac, B., Lert, F., Rouillon, F., Bénichou, J., et al. (2011). Benchmarking clinical management of spinal and non-spinal disorders using quality of life: results from the EPI3-LASER survey in primary care. *European Spine Journal*, 20:2210-2216.

Rothwell, P.M., McDowell, Z., Wong, C.K., Dorman, P.J. (1997). Doctors and patients don't agree: cross-sectional survey of patients' and doctors' perceptions and disability in multiple sclerosis. *British Medical Journal*, 314:1580-1583.

Rowbotham, D.J., MacIntyre, P.E. (2003) *Clinical Pain Management: Acute Pain*. London: Arnold.

Royal College of General Practitioners (1999). *Management of Acute Low Back Pain*. London: Royal College of General Practitioners.

Rubin, H.R. (1990) Can patients evaluate the quality of hospital care? *Medical Care Review* 47:267-325.

Rubinstein, L.V., Calkins, D.R., Young, R.T., Cleary, P.D., Fink, A., Kosecoff, J. (1989). Improving patient function: A randomised controlled trial of functional disability screening. *Annals of Internal Medicine*, 111(10): 836-842.

Rudd, P. (1997) In search of the gold standard for compliance measurement. *Archives of Internal Medicine* 139:627-8.

Rugg, S., Paterson, C., Britten, N., Bridges, J., Griffiths, P. (2011) Traditional acupuncture for people with medically unexplained symptoms: a longitudinal qualitative study of patients' experiences. *Br J Gen Pract*. 61(587):e306-15.

Russek, L., Wooden, M., Ekedahl, S., Bush, A. (1997) Attitudes towards standardized data collection. *Physical Therapy* 77(7):714-729.

Russell, T.G., Wootton, R., Jull, G.A. (2002) Physical outcome measurements via the Internet: reliability at two Internet speeds. *Journal of Telemedicine and Telecare* 8 Suppl 3:S3:50-2.

Ryan, C.M., Lee, A.F., Kazis, L.E., Shapiro, G.D., Schneider, J.C., Goverman, G., Fagan, S.P., Wang, C., Kim, J., Sheriden, R.L., Tompkins, R.G. (2016) Is real-time feedback of burn-specific patient-reported outcome measures in clinical settings practical and useful? A pilot study implementing the Young Adult Burn Outcome Questionnaire. *Journal of Burn Care and Research* 37(1):64-74.

Rychetnick, I., Frommer, M., Hawe, P., Shiell, A. (2002) Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health* 56:119-127.

Rycroft-Malone, J., Bucknall, T. (2010) Using theory and frameworks to facilitate the implementation of evidence into practice. *Worldviews Evid Based Nurs*. 7(2):57-8.
Sabino, J., Grauer, J.N. (2008) Pregnancy and low back pain. *Current Reviews in Musculoskeletal Medicine* 1**,**137-41.

SAC (Scientific Advisory Committee) of the Medical Outcomes Trust (2002). Instrument review criteria. Assessing health status and quality of life instruments: attributes and review criteria. *Quality of Life Research* 11:193-205.

Sackett, D. (2006). Haynes, R. Brian, ed. *Clinical Epidemiology: How to Do Clinical Practice Research*. Lippincott Williams & Wilkins.

Salaffi, F., Gasparini, W., Grassi, W. (2009) The use of computer touch-screen technology for the collection of patient-reported outcome data in rheumatoid arthritis: comparison with standardized paper questionnaires. *Clinical and Experimental Rheumatology* 27: 459-468.

Saleh, K.J., Radsosevich, D.M., Kassim, R.A. (2002) Comparison of commonly used orthopaedic outcome measures using palm-top computers and paper surveys. *Journal of Orthopaedic Research* 20:1146-1151.

Salisbury, C., Johnson, L., Purdy, S., Valderas, J.M., Montgomery, A.A. (2011) Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *Br J Gen Practice*. 61(582): e12–e21.

Salisbury, C., Wallace, M., Montgomery, A.A. (2010) Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling. *BMJ.* 341:c5004

Sammut-Bonnici, T., Galea, F. "PEST Analysis" in *Wiley Encyclopedia of Management,* John Wiley, 2015, London.

Santana, M-J., Feeny, D. (2014). Framework to assess the effects of using patient-reported outcome *Quality of Life Research* measures in chronic care management. , 23:1505–1513.

Saper, N. (2013). International Cryptography Regulation and the Global Information Economy. *North Western Journal of Technology and Intellectual Property* 11(7):673-688.

Saranto, K., Kinnunen, U-M. (2009) Evaluating nursing documentation – research designs and methods: systematic review. *Journal of Advanced Nursing* 65(3):464-476.

Sassen, B., Kok, G., Vanhees, L. (2011) Predictors of healthcare professionals' intention and behaviour to encourage physical activity in patients with cardiovascular risk factors. *BMC Public Health* 11:246.

Savigny, P., Kuntze, S., Watson, P., Underwood, M., Ritchie, G., Cotterell, M., Hill, D., Browne, N., Buchanan, E., Coffey, P., Dixon, P., Drummond, C., Flanagan, M., Greenough, C., Griffiths, M., Halliday-Bell, J., Hettinga, D., Vogel, S., Walsh, D. (2009). *Low back pain: early management of persistent non-specific low back pain.* London: National Collaborating Centre for Primary Care and Royal College of General Practitioners.

Sawyer, C. E., Evans, R.L., Boline, P.D., Branson, R., Spicer, A. (1999). A feasibility study of chiropractic spinal manipulation versus sham spinal manipulation for chronic otitis media with effusion in children. *J Manipulative Physiol Ther*. 22(5): 292-298.

Scally, G., Donaldson, L.J. (1998) *The NHS 's 50th anniversary*. *Clinical governance and the drive for quality improvement in the NHS in England*. London: Department of Health.

Schaefer, D.R., Dillman, D.A. (1998) Development of a standard e-mail methodology. *Public Opinion Quarterly* 62:378-97.

Shafrir, Y. and B. A. Kaufman (1992). Quadriplegia after chiropractic manipulation in an infant with congenital torticollis caused by a spinal cord astrocytoma. *J Pediatr*. 120(2 Pt 1): 266-269.

Schapira, M.M., Nattinger, A.B., McAuliffe, T.L. (2006) The influence of graphic format on breast cancer risk communication. *Journal of Health Communication* 11:569-582.

Schellingerhout, J.M., Verhagen, A.P., Heymans, M.W., Koes, B.W., de Vet, H.C., Terwee, C.B. (2012) Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res.* 21(4):659-70.

Schlessinger, M., Grob, R., Shaller, D. (2016). Using patient-reported information to improve clinical practice. *Health Services Research*, 50(S2): 2116-2154.

Schmidt, H. (ed) (1984) *Tutorials in problem-based learning*. Maastricht: Van Gorcum.

Schmitt, M.A., de Wijer, A., van Genderen, F.R. (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* (Phila Pa 1976). 34(23):2551-61.

Schmitt, M.A., Schröder, C.D., Stenneberg, M.S., van Meeteren, N.L., Helders, P.J., Pollard, B., Dixon, D. (2013) Content validity of the Dutch version of the Neck Bournemouth Questionnaire. *Man Ther.* 18(5):386-9.

Scientific Advisory Committee of the Medical Outcomes Trust (1995). Instrument review criteria. *Medical Outcomes Trust Bulletin* 3(4):I-IV.

Scott, J., Huskisson, E.C. (1976). Graphic representation of pain. *Pain* 2(2):175-184.

Scott, J., Huskisson, E.C. (1979) Accuracy of subjective measurements made with or without previous scores: an important source of error in serial measurement of subjective states. *Annals of Rheumatic Diseases* 38:558-559.

Seers, K., Cox, K., Crichton, N.J., Edwards, R.T., Eldh, A.C., Estabrooks, C.A., Harvey, G., Hawkes, C., Kitson, A., Linck, P., McCarthy, G., McCormack, B., Mockford, C., Rycroft-Malone, J., Titchen, A., Wallin, L. (2012) FIRE (Facilitating Implementation of Research Evidence): a study protocol. Implement Sci. 7:25.

Semmelweis, I.P. (1860) A gyermekágyi láz fölötti véleménykülönbség köztem s az angol orvosok közt. *Orvosi hetilap* 849–851, 873-76, 889–893, 913–915.

Senstad, O., Leboeuf-Yde, C., Borchgrevink, C. (1996). Predictors of side effects to spinal manipulative therapy. *J Manipulative Physiol Ther*. 19(7): 441-445.

Senstad, O., Leboeuf-Yde, C., Borchgrevink, C. (1997). Frequency and characteristics of side effects of spinal manipulative therapy. *Spine* (Phila Pa 22(4): 435-440.

Seymour, R.A. The use of pain scales in assessing the efficacy of analgesics in post-operative pain. *European Journal of Clinical Pharmacology*. 23:441-444, 1982.

Serlin, R.C., Mendoza, T.R., Nakamura, Y., Edwards, K.R., Cleeland, C.S. (1995) When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain* 61:277-284.

Seymour, R.A. (1982) The use of pain scales in assessing the efficacy of analgesics in post-operative dental pain. *European Journal of Clinical Pharmacology* 23(5):441-444.

Shah, P., Mountain, D. (2007) The medical model is dead – long live the medical model. *British Journal of Psychiatry* 191:375-377.

Shekelle, P.G., Andersson, G., Bombardier, C., Cherkin, D., Deyo, R., Keller, R., Lee, C., Liang, M., Lipscomb, B., Spratt, K. (1994) A brief introduction to the critical reading of the clinical literature. *Spine* (Phila Pa 1976)19(18 Suppl):2028S-2031S.

Sherman, K.J., Cherkin, D.C., Ichikawa, L., Avins, A.L., Delaney, K., Barlow, W.E., Khalsa, P.S., Deyo, R.A. (2010) Treatment expectations and preferences as predictors of outcome of acupuncture for chronic back pain. *Spine* (Phila Pa 1976). 35(15):1471-7.

Shih, T.H., Fan, X. (2008) Comparing response rates from web and mail surveys: a meta-analysis. *Field Methods* 20(3): 249–271.

Shuy, R. W. (2003). In-person versus telephone interviewing. In J. A. Holstein & J. F. Gubrium (Eds), *Inside interviewing: New lenses, new concerns* Thousand Oaks: Sage.

Shrout, P.E., Fleiss, J.L. (1997) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 86(2):420-8.

Silverman, D. (2011) *Interpreting qualitative data*. Fourth edition. Thousand Oaks, CA, US: Sage Publications.

Silverman, S.L. (2009) From randomized controlled trials to observational studies. *American Journal of Medicine* 122(2):114-120.

Sitzia, J., Wood, N. (1997) Patient satisfaction: a review of issues and concepts. *Soc Sci Med*. 45(12):1829-43.

Skeat, J., Perry, A. (2008) Exploring the implementation and use of outcome measurement in practice: a qualitative study. *International Journal of Language and Communication Disorders* 43(2):110-125.

Somner, J.E., Sii, F., Bourne, R.R., Cross, V., Burr, J.M., Shah, P. (2012) Moving from PROMs to POEMs for glaucoma care: a qualitative scoping exercise. *Invest Ophthalmol Vis Sci.* 53(9):5940-7.

Smeets, R., Köke, A., Lin, C-W., Ferreira, M., Demoulin, C. (2011) Measures of function in low back pain/disorders: Low Back Pain Rating Scale (LBPRS), Oswestry Disability Index (ODI), Progressive Isoinertial Lifting Evaluation (PILE), Quebec Back Pain Disability Scale (QBPDS), and Roland-Morris Disability Questionnaire (RDQ). *Arthritis Care Research* 63(S11):S158-S173.

Smith, B.H., Penny, K.I., Purves, A.M. (1997) The Chronic Pain Grade questionnaire: validation and reliability in postal research. *Pain* 71:141-7.

Smith, H.J., Dinev, T., Xu, H. (2011) Information privacy research: an interdisciplinary review. *Management Information System Quarterly*. 11;35:989-1016

Smith, R., Rennie, D. (2014) Evidence based medicine-an oral history. *British Medical Journal* 348:g371.

Smith, S.C., Cano, S., Lamping, D.L., Staniszewska, S., Browne, J., Lewsey, J., van der Meulen, J., Cairns, J., Black, N. (2005) *Patient Reported Outcome Measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific evidence*. London: Department of Health.

Snape, D., Kirkham, J., Britten, N., Froggatt, K., Gradinger, F., Lobban, F., Popay, J., Wyatt, K., Jacoby, A. (2014) Exploring perceived barriers, drivers, impacts and the need for evaluation of public involvement in health and social care research: a modified Delphi study. *BMJ Open* 4(6):e004943

Snelling, N. J. (2006) Spinal manipulation in patients with disc herniation: a critical review of risk and benefit. *International Journal of Osteopathic Medicine.* 9(3): 77-84.

Snow, S., Kirwan, J.R. (1988) Visual analogue scales: a source of error. *Annals of the Rheumatic Diseases* 47:526.

Snyder, C.F., Jensen, R.E., Segal, J.B., Wu, A.W. (2013) Patient reported outcomes (PROs): putting the patient perspective in patient-centred outcomes research. *Med Care.* 51(8 0 3): S73–S79.

Sodha, R., Sivanadarajah, N., Alam, M. (2013). The use of glucosamine for chronic low back pain: a systematic review of randomised control trials. *British Medical Journal Open* 3(6).

Soklic, M., Peterson, C., Humphreys, B.K. (2012) Translation and validation of the German version of the Bournemouth Questionnaire for Neck Pain. *Chiropractic & Manual Therapies* 20:2.

Somerville, S., Hay, E., Lewis, M., Barber, J., van der Windt, D., Hill, J., Sowden, G. (2008). Content and outcome of usual primary care for back pain: a systematic review. *British Journal of General Practice* 58(556):790-7, i-vi.

Sørensen, H.T., Lash, T.L., Rothman, K.J. (2006) Beyond Randomized Controlled Trials: A Critical Comparison of Trials With Nonrandomized Studies. *Hepatology* 44:1075-1082.

Soton, 2016. Introduction to Psychological Theories. Retrieved 13th May, 2016 from http://www.southampton.ac.uk/psychology/postgraduate/research_students/mmh1e13.page.

Spanjer, J., Groothoff, J.W., Brouwer, S. Instruments used to assess functional limitations in workers applying for disability benefit: a systematic review. *Disability and Rehabilitation* 33(23-24):2143-50.

Sperry, K., Pfalzgraf, R. (1990). Inadvertent clavicular fractures caused by "chiropractic" manipulations in an infant: an unusual form of pseudoabuse. *J Forensic Sci*. 35(5): 1211-1216.

Squires, J.E., Grimshaw, J.M., Taljaard, M., Linklater, S., Chassé, M., Shemie, S.D., Knoll, G.A. (2014) Design, implementation, and evaluation of a knowledge translation intervention to increase organ donation after cardiocirculatory death in Canada: a study protocol. *Implement Sci*. 9:80.

Standard Contracts for Acute NHS Services (2008). London: Department of Health.

Stame, N. (2002). *Evaluation in Italy: an inverted sequence from performance management to programme evaluation*, In Furabo, J.C., Rist, R.C., Sandahl, R. (eds) *International Atlas of Evaluation*. New Brunswick and London: Transaction (273-290)

Stanley, I., al-Shehri, A., Thomas, P. (1993) Continuing education for general practice. Experience, competence and the media of self-directed learning for established general practitioners. *British Journal of General Practice* 43(370):210-4.

Starfield, B. (1979) Measuring the attainment of primary care. *Journal of Medical Education* 54(5):361-369.

Starfield, B. (2005) Measurement of outcome: a proposed scheme. *Millbank Quarterly.* 83:1-11.

Statista.com. (2016) Estimated number of apps available for each operating system. Retrieved 30th June, 2016 from http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/.

Steele, K. M., Carreiro, J.E., Viola, J.H., Conte, J.A., Ridpath, L.C. (2014). Effect of osteopathic manipulative treatment on middle ear effusion following acute otitis media in young children: a pilot study. *J Am Osteopath Assoc*. 114(6): 436-447.

Steinhubl, S.R., Muse, E.D., Topol, E.J. (2013) Can mobile health technologies transform health care? *Journal of the American Medical Association* 310:2395-6.

Stephenson, J.M., Strange, V., Forrest, S., Oakley, A., Copas, A., Allen, E., Babiker, A., Black, S., Ali, M., Monteiro, H., Johnson, A.M.; RIPPLE study team. Pupil-led sex education in England (RIPPLE study): cluster-randomised intervention trial. *Lancet.* 364(9431):338-46.

Stevens, K., Palfreyman, S. (2012) The use of qualitative methods in developing the descriptive systems of preference-based measures of health-related quality of life for use in economic evaluation. *Value Health* 15(8):991-8.

Still AT. *Osteopathy Research and Practice*. (1992) Seattle, Walsh: Eastland Press.

Stimson, G.V. (1974). Obeying doctors' orders: a view from the other side. *Social Science and Medicine*, 8(2):97-104.

Stone, A.A., Broderick, J.E., Schwartz, J.E., Shiffman, S., Litcher-Kelly, L., Calvanese, P. (2003) Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain.* 104(1-2):343-51.

Stratford, P.W., Binkley, J., Solomon, P. (1994) Assessing change over time in patients with low back pain. *Physical Therapy* 74(6):528-533.

Stratford, P.W., Binkley, J., Solomon, P., Finch, E., Gill, C., Moreland, J. (1996a) Defining the minimum detectable change for the Roland-Morris Disability Questionnaire. *Physical Therapy* 76(4): 359-365.

Stratford, P.W., Finch, E., Solomon, P., Binkley, J., Gill, C., Moreland, J. (1996b) Using the Roland Morris Questionnaire to make decisions about individual patients. *Physiotherapy Canada* 48:107-110.

Stratford, P.W., Binkley, J.M. (1997) Measurement properties of the RM-18. A modified version of the Roland Morris Disability Scale. *Spine* 22(20):2416-2421.

Stratford, P.W., Binkley, J., Riddle, D.L., Guyatt, G.H. (1998) Sensitivity to change of the Roland Morris Back Pain Questionnaire: Part I: *Physical Therapy* 78(11):1186-1196.

Stratford, P.W., Binkley, J.M. (1999) Applying the results of self-report measures to individual patients: an example using the Roland-Morris Questionnaire. *Journal of Orthopaedics and Sports Physical Therapy* 29(4):232-239.

Stratford, P.W., Riddle, D.L. (2005). Assessing sensitivity to change: choosing the appropriate change coefficient. *Health Quality of Life Outcomes* 3:23.

Stratford, P.W., Kennedy, D.M., Woodhouse, L.J., Spadoni, G.F. (2007). Measurement properties of the WOMAC LK 3.1 pain scale. *Osteoarthritis and Cartilage*, 15:266-272.

Strauss, A., Corbin, J. (1998) *Basics of qualitative research: Techniques and procedures for developing grounded theory*, 2nd ed. Thousand Oaks, CA, US: Sage Publications,

Streiner, D.L, Norman, G.R. (1995) *Health measurement scales*. 2nd ed. Oxford: Oxford University Press.

Streiner, D.L. (2003) Clinimetrics vs. psychometrics: an unnecessary distinction. *Journal of Clinical Epidemiology* 56(12):1142-5.

Stroud, M.W., McKnight, P.E., Jensen, M.P. (2004) Assessment of self-reported physical activity in patients with chronic pain: development of an abbreviated Roland-Morris Disability Scale. *Journal of Pain* 5:257-263.

Strutt, R., Shaw, Q., Leach, J. (2008) Patients' perceptions and satisfaction with treatment in a UK osteopathic training clinic. *Manual Therapy* 13(5):456-67.

Stupply, D.J. (1998) The Faces Pain Scale: reliability and validity with mature adults. *Applied Nursing Research* 11:84-9.

Sturges, J.E., Hanrahan, K.J. (2004) Comparing telephone and face-to-face qualitative interviewing: A research note. *Qualitative Research* 4:107–118.

Svensson, E. (2001) Guidelines to statistical evaluation of data from rating scales and questionnaires. *J Rehabil Med*. 33:47–8.

Swarbrick, G. (2013). Personal communication with Mr Swarbrick, Head of Healthcare Outcomes at BUPA Healthcare.

Sweeney, G., O'Hagan, B., Squire, S., Powell, C. (2005) The Patients Accelerating Change Project: does it make any difference? *Clinical Governance* 10(1):72 –83.

Swinkels, R.A., van Peppen, R.P., Wittink, H., Custers, J.W., Beurskens, A.J. (2011) Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. *BMC Musculoskeletal Disorders* 12:106.

Tadić, V., Knowles, R.L., Rahi, J. (2013) Data from the Multi-professional Workshop: Paediatric Patient-Reported Outcome and Experience Measures (PROMS and PREMS) in Routine Clinical Practice. *Journal of Open Public Health Data* 1(1):e4

Tarsuslu, T., Bol, H., Simşek, I.E., Toylan, I.E., Cam, S.. (2009). The effects of osteopathic treatment on constipation in children with cerebral palsy: a pilot study. *J Manipulative Physiol Ther.* 32(8): 648-653.

Tausig, J.E., Freeman, E.W. (1988) The Next Best Thing to Being There: Conducting the Clinical Research Interview by Telephone. *American Journal of Orthopsychiatry* 58(3): 418–27.

Tavabie, J.A., Tavabie, O.D. (2009). Improving care in depression: qualitative study investigating the effects of using a mental health questionnaire. *Quality in Primary Care*, 17(4):251-261.

Taylor, A.L, (1984). The Wizard Inside the Machine. *TIME* 123(16):56-63.

Taylor, D., Bury, M., Campling, N., Carter, S., Garfield, S., Newbould, J., Rennie, T. (2006) A review of the use of the health belief model (HBM), the theory of reasoned action (TRA), the theory of planned behaviour (TPB), and the trans-theoretical model (TTM) to study and predict health related behaviour change. National Institute for Health and Care Excellence. Retrieved 12th February, 2016 from https://www.nice.org.uk/.../behaviour-change-taylor-et-al-models-review2 .

Taylor, D.C., Hamdy, H. (2013) Adult learning theories: implications for learning and teaching in medical education: AMEE Guide No. 83. *Medical Teacher* 35(11):e1561-72.

Taylor, S.J., Carnes, D., Homer, K., Kahan, B.C., Hounsome, N., Eldridge, S., Spencer, A., Pincus, T., Rahman, A., Underwood, M. (2016) Novel Three-Day, Community-Based,

Nonpharmacological Group Intervention for Chronic Musculoskeletal Pain (COPERS): A Randomised Clinical Trial. *PLoS Medicine* 13(6):e1002040.

Terwee, C.B., Dekker, F.W., Wiersinga, W.M., Prummel, M.F., Bossuyt, P.M. (2003) On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research* 12: 349-362.

Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M., de Vet, H.C. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 60(1):34-42.

Terwee, C.B., Jansma, E.P., Riphagen, I.I., de Vet, H.C. (2009a) Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009a 18(8):1115-23.

Terwee, C.B., Roorda, L.D., Knol, D.L., De Boer, M.R., De Vet, H.C. (2009b) Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol.* 62(10):1062-7.

Terwee, C.B., Mokkink, L.B., Knol, D.L., Ostelo, R.W., Bouter, L.M., de Vet, H.C. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* 21(4):651-7.

Terwee, C.B., de Vet, H.C.W., Prinsen, C.A.C., Mokkink, L.B. (2016). Comment on "Checklist to operationalize measurement characteristics of patient-reported outcome measures" http://www.cosmin.nl/30/comment-on-article-by-francis-et-al

Testa, M. (1987) Interpreting quality of life clinical trial data for use in the clinical practice of antihypertensive therapy. *Journal of Hypertension* Suppl 5:S9-S13.

The AGREE Collaboration (2003). Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality and Safety in Health Care* 12:18–23.

Thomas, K.J., Nicholl, J.P., Coleman, P. (2001) Use and expenditure on complementary medicine in England: a population based survey. *Complementary Therapies in Medicine* 9:2-11.

Thomas, K.J., Coleman, P., Nicholl, J.P. (2003) Trends in access to complementary or alternative medicines via primary care in England: 1995-2001 results from a follow-up national survey. *Family Practice* 20(5):575-7.

Thompson, A.G.H. (2007) The meaning of patient involvement and participation in health care consultations: a taxonomy. *Social Science and Medicine* 64(6):1297-1310.

Thompson ,A., Sunol, R. (1995) Expectations as determinants of patient satisfaction: concepts, theory and evidence. *International Journal for Quality in Health Care* 7(2): 127-141

Thomson, Bruce S. (2004). *Qualitative research: Grounded theory—Sample size and validity* Retrieved 8th September, 2009 from http://www.buseco.monash.edu.au/research/studentdocs/mgt.pdf .

Thomson, P., Walker, M. (eds) (2010) *The Routledge Doctoral Student's Companion: Getting to Grips with Research in Education and the Social Sciences (Companions for PhD and DPhil Research).* Oxford: Routledge.

Tiplady, B., Crompton, G.K., Brackenridge, D. (1995) Electronic diaries for asthma. *BMJ.* 310(6992):1469.

Todd, A. J., Carroll, M.T., Robinson, A., Mitchell, E.K. (2015). Adverse Events Due to Chiropractic and Other Manual Therapies for Infants and Children: A Review of the Literature. *J Manipulative Physiol Ther*. 38(9): 699-712.

Tong, A., Sainsbury, P., Craig, J. (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32- item checklist for interviews and focus groups. *International Journal of Quality in Health Care*19(6):349-357.

Toomey, T.C., Seville, J.L., Mann, J.D. (1995) The pain locus of control scale: relationship to pain description, self-control skills and psychological symptoms. *The Pain Clinic* 8:315-22.

Trauer, T., Gill, L., Pedwell, G., Slattery, P. (2006) Routine outcome measurement in public mental health – what do clinicians think? *Australian Health Review* 30(2):144-147.

Tubach, F., Ravaud, P., Baron, G., Falissard, B., Logeart, I., Bellamy, N. (2005) Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheumat Dis*. 64:29–33.

Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem*. Proceedings of the London Mathematical Society. 42 (1): 230–265.*

Turk, D.C., Melzack, R. (1992) The measurement and assessment of people in pain. In *Handbook of Pain Assessment* (Turk DC and Melzack R eds). New York: The Guildford Press.

Uhlmann, R.E., Inui, T.S., Carter, W.B. (1984) Patient requests and expectations. Definitions and clinical applications. *Medical Care* 22:681-5.

UK BEAM trial team. (2004a) United Kingdom Back Pain, Exercise and Manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *British Medical Journal* 329:1377.

UK BEAM trial team (2004b). United Kingdom Back Pain, Exercise and Manipulation (UK BEAM) randomised trial: cost effectiveness of physical treatments for back pain in primary care. *British Medical Journal* 329:1381.

UK Government, 2010. Government policy on NHS efficiency. Retrieved 21st July, 2016 from https://www.gov.uk/government/publications/2010-to-2015-government-policy-nhs-efficiency.

Underwood, M.R., Barnett, A.G., Vickers, M.R. (1999) Evaluation of two time-specific back pain outcome measures. *Spine* 24(11):1104-1112.

Underwood, M., Harding, G., and Klaber Moffett, J. (2006) Patient perceptions of physical therapy within a trial for back pain treatments (UK BEAM.) *Rheumatology* 45:751–756.

Underwood, M., Parsons, S., Eldridge, S., Spencer, A., Feder, G. (2006) Asking older people about fear of falling did not have a negative event. *Journal of Clinical Epidemiology* 59(6):629-634.

Valderas, J.M., Alonso, J., Prieto, L., Espallargues, M., Castells, X. (2004). Content-based interpretation aids for health-related quality of life measures in clinical practice. An example for the visual function index (VF-14). *Quality of Life Research*, 13(1):35-44.

Valderas, J.M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C.E., Halyard, M.D. (2008a) The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Quality of Life Research* 17:179-93.

Valderas, J.M., Ferrer, M., Mendívil, J., Garin, O., Rajmil, L., Herdman, M., Alonso, J.; Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network (2008b). Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health* 11(4):700-8.

Valderas, J.M., Fitzpatrick, R., Roland, R., M. (2012) Using health status to measure NHS performance: another step in the dark for health reform in England. *BMJ Quality and Safety* 21(4):352-353.

Valier, A.R., Jennings, A.L., Parsons, J.T., Vela, L.I. (2014) Benefits of and barriers to using patient-rated outcome measures in athletic training. *Journal of Athletic Training* 49(5):674-83.

van der Roer, N., Ostelo, R.W., Bekkering, G.E., van Tulder, M.W., de Vet, H.C. (2006) Minimal clinically change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine* (Phila Pa 1976) 31(5):578-82.

Van der Wees, P.J., Nijhuis- Van der Sanden, M.W.G., Ayanian, J.Z., Black, N., Westert, G.P., Schneider, E.C. (2014). Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *The Milbank Quarterly* 92(4):754-775.

Van Noorden, R. (2016) Global scientific output doubles every nine years. http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html

Van Peppen, R.P.S., Maissan, F.J.F., van Genderen, F.R., van Dolder, E., Meeteren, N.L.U. (2008) Outcome measures in physiotherapy management of patients with stroke: a survey into self-reported use, and barriers to, and facilitators for use. *Physiotherapy Research Internationa*l 123(4):255-270.

van Poecke, A. J., Cunliffe, C. (2009). Chiropractic treatment for primary nocturnal enuresis: a case series of 33 consecutive patients. *J Manipulative Physiol Ther*. 32(8): 675-681.

Van Tubergen, A., Debats, I., Ryser, L., Londoño, J., Burgos-Vargas, R., Cardiel, M.H., Landewé, R., Stucki, G., Van Der Heijde, D. (2002) Use of a numerical rating scale as an answer modality in ankylosing spondylitis-specific questionnaires. *Arthritis and Rheumatism* 47(3):242-8.

Van Woerkom, C., Adolfse, L. (1998) Interactive development and use of knowledge. *SI* 1:10-19.

Vavrek, D., Haas, M., Neradilek, M.B., Polissar, N. (2015). Prediction of pain outcomes in a randomised controlled trial of dose-response of spinal manipulation for the care of chronic low back pain. *BMC Musculoskeletal Disorders*, 16:205-218.

Velicova, G., Booth, L., Smith, A.B. (2004) Measuring quality of life in routine oncology practice improves communication and patient wellbeing: a randomized controlled trial. *Journal of Clinical Oncology* 22(4):714-724.

Versi, E."Gold standard"is an appropriate term. *British Medical Journal* 305:187.

Vogel, S. Clinical Risks Osteopathy and Management (CROaM) full report. (2009) Retrieved 1st May, 2014 from http://www.osteopathy.org.uk/uploads/croam_full_report_0313.pdf.

Vohra, S., Johnston, B.C., Cramer, K., Humphreys, K. (2007). Adverse events associated with pediatric spinal manipulation: a systematic review. *Pediatrics.* 119(1): e275-283.

Vohra, S., Brulotte, J., Le, C., Charrois, T., Laeeque, H. (2009). Adverse events associated with paediatric use of complementary and alternative medicine: Results of a Canadian Paediatric Surveillance Program survey. *Paediatr Child Health*. 14(6): 385-387.

Von Korff, M., Saunders, K. (1996) The course of back pain in primary care. *Spine* 21:2833-2837.

Von Korff, M., Jensen, M.P., Karoly, P. (2000) Assessing global pain severity by self-report in clinical and health services research. *Spine* 25:3140-3151.

Waddell, G., Newton, M., Henderson, I., Somerville, D., Main, C.J. (1993) A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 52:157-68.

Wagner, A.K., Ehrenberg, B.L., Tran, T.A., Bungay, K.M., Cynn, D.J., Rogers, W.H. (1997). Patient-based health status measurement in clinical practice: A study of its impact on epilepsy patients' care. *Quality of Life Research*, 6:329-341.

Wahl, R. A., Aldous, M.B., Worden, K.A., Grant, K.L. (2008). Echinacea purpurea and osteopathic manipulative treatment in children with recurrent otitis media: a randomized controlled trial. *BMC Complement Altern Med*. 8:56.

Walker, B.F. (1998) The prevalence of low back pain: a systematic review of the literature from 1966 to 1998. *Journal of Spinal Disorders* 13 (3):205-217.

Waller, P., Evans, S. (2011) Mediator: who's to blame? *Lancet*. 11;377(9782):2002-3.

Walport, M. (2014). Uses and Abuses of Patient Reported Outcome Measures (PROMs): Potential Iatrogenic Impact of PROMs Implementation and How It Can Be Mitigated. *Adm Policy Ment Health,* 41:141–145.

Walsh, D., Downe, S. (2005) Meta-synthesis method for qualitative research: a literature review. *JAN* 50(2): 204-211.

Walters, S.J., Brazier, J.E. (2003). What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and Quality of Life Outcomes* 1:4.

Wang, D., Tsui, A.S., Zhang, Y., & Ma, L. (2003). Employment relationships and firm performance: evidence from an emerging economy. *Journal of Organizational Behavior*, 24, 511-535.

Wang, W.L., Lee, H.L., Fetzer, S.J. (2006) Challenges and strategies of instrument translation. *Western Journal of Nursing Research* 28:310-321.

Wang, Y., Zhu, W.L., Dong, Y.F. (2008). Massage manipulation of supplementing marrow and kneading tendon in treating 30 children with spastic cerebral palsy. *Zhongguo Zhong Xi Yi Jie He Za Zhi.* 28(4): 363-365.

Ware, J.E., Sherbourne, C.D. (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 30: 473–483.

Ware, J.E., Snyder, M.K. (1975) Dimensions of patient attitudes regarding doctors and medical care services. *Medical Care* 13:669-682.

Weber, M. 1946/1958. *Essays in Sociology*. (H. Gerth and C. Mills, Eds. And Trans.) In M. Weber, H. Gerth, & C. Mills, From Max Weber. New York: Oxford University Press.

Weber, M. (1991) *The Nature of Social Action in Runciman*, W.G. *Weber: Selections in Translation* Cambridge: Cambridge University Press.

Weber Hellstenius, S. A. (2009). Recurrent neck pain and headaches in preadolescents associated with mechanical dysfunction of the cervical spine: a cross-sectional observational study with 131 students. *J Manipulative Physiol Ther.* 32(8):625-34.

Weiner, B.J. (2009) A theory of organizational readiness for change. *Implement Sci.* 4:67.

Wells, G., Beaton, D., Shea, B., Boers, M., Simon, L., Strand, V., Brooks, P., Tugwell, P. (2001) Minimal clinically important differences: review of methods. *Journal of Rheumatology* 28: 406-412.

Wenig, C.M., Schmidt, C.O., Kohlmann, T., Schweikert, B. (2009) Costs of back pain in Germany. European Journal of Pain 13(3):280-6.

Wennberg, J.E., Barry, M.J., Fowler, F.J. (1993) Outcomes research, PORTs, and Health Care reform. *Annals of New York Academy of Sciences*. 52-62.

Wensing, M., Grol, R., Smits, A. (1994) Quality judgements by patients on general practice care: a literature analysis. Social Science in Medicine 38:45-53.

Wensing, M. (2000) Evidence-based patient empowerment. *Quality in Health Care* 9:200-1.

Wensing, M., Elwyn, G. (2003) Methods for incorporating patients' views in health care. *British Medical Journal* 326:877-879.

Wiberg, J.M., Nordsteen, J., Nilsson, N. (1999) The short-term effect of spinal manipulation in the treatment of infantile colic: a randomized controlled clinical trial with a blinded observer. *J Manipulative Physiol Ther*. 22(8):517-22.

Wight, D., Raab, G.M., Henderson, M., Abraham, C., Buston, K., Hart, G., Scott, S. (2002) Limits of teacher delivered sex education: interim behavioural outcomes from randomised trial. *BMJ.* 324(7351):1430.

Wiles, R., Cott, C., Gibson, B. (2009) Hope, expectations and recovery from illness: a narrative synthesis. *Journal of Advanced Nursing*. 64(6):564-573.

Wilkinson, J., Peters, D., Donaldson, J. (2004). *Clinical Governance for Complementary and Alternative Medicine in Primary Care*. London: University of Westminster.

Wilkinson, S. (1998) Focus groups in feminist research: Power, interaction, and the co-construction of meaning. *Women's Studies International Forum*. 21(1):111-125.

Williams, B.A., Kentor, M.L., Vogt, M.T., Irrgang, J.J., Bottegal, M.T., West, R.V., Harner, C.D., Fu, F.H., Williams, J.P. (2006) Reduction of verbal pain scores after anterior cruciate ligament reconstruction with 2-day continuous femoral nerve block: a randomized clinical trial. *Anesthesiology*. 104(2):315-27.

Williams, D.P., Price, A.J., Beard, D.J. (2013) The effects of age on patient-reported outcome measures in total knee replacements. *Bone and Joint Journal* 95-B(1):38-44.

Williams, N.H., Wilkinson, C., Russell, I. (2003) Randomized osteopathic manipulation study. ROMANS): pragmatic trial for spinal pain in primary care. *Family Practice* 20**,** 662-9.

Williams, N.H., Hendry, M., Lewis, R. (2007) Psychological response in spinal manipulation (PRISM): a systematic review of psychological outcomes in randomised controlled trials. *Complementary Therapies in Medicine* 15(4):271-83.

Williams, R.M., Myers, A.M. (2001) Support for a shortened Roland-Morris Disability Questionnaire for patients with acute low back pain. *Physiotherapy Canada* 53: 60-66.

Williams, S.J., Calnan, M. (1991) Key determinants of consumer satisfaction with general practice. *Family Practice* 8(3):237-42.

Williams, S., Weinman, J., Dale J. (1995) Patient expectations: What do primary care patients want from the GP and how far does meeting expectations affect patient satisfaction? *Family Practice.* 12(2): 193-201

Williamson, A., Hoggart, B. (2005) Pain: a review of three commonly used pain rating scales. *Journal of Clinical Nursing* 14:798-804.

Wippert, P-M., Fliesser, M., Krause, M. (2017). Risk and protective factors in the clinical rehabilitation of chronic back pain. *Journal of Pain Research*, 10:1569-1579.

Wong,D.L., Baker, C.M. (1988) Pain in children: comparison of assessment scales. *Pediatr Nurs.* 14(1):9-17.

Woo, C. C. (1993). Post-traumatic myelopathy following flopping high jump: a pilot case of spinal manipulation. *J Manipulative Physiol Ther*. 16(5): 336-341.

World Health Organisation Scientific Group. (2003) The Burden of Musculoskeletal Conditions at the start of the new Millenium. WHO Scientific Group Technical Report Series 919. www.whqlibdoc.who.int/trs/WHO_TRS_919.pdf

Wressle, E., Lindstrand, J., Neher, M. *Disability and Rehabilitation* 25(10):497-506.

Wright, J.G. (2000) Evaluating the outcome of treatment : shouldn't we be asking patients if they are better? *Journal of Clinical Epidemiology* 53:549-553.

Wright, J.G., Young, N.L. (1997) A comparison of different indices of responsiveness. *Journal of Clinical Epidemiology* 50: 239-246.

Wyatt, K., Edwards, V., Franck, L., Britten, N., Creanor, S., Maddick, A., Logan, S. (2011). Cranial osteopathy for children with cerebral palsy: a randomised controlled trial. *Arch Dis Child*. 96(6): 505-512

Wye, L., Sharp, D., Shaw, A. (2009) The impact of NHS based primary care complementary therapy services on health outcomes and NHS costs: a review of service audits and evaluations. *BMC Complementary and Alternative Medicine* 9**:**5.

Wyrwich, K.W., Tierney, W.M., Wolinsky, F.D. (1999a) Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol*. 52(9):861-73.

Wyrwich, K.W., Nienaber, N.A., Tierney, W.M., Wolinsky, F.D. (1999b) Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care*. 37(5):469-78.

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

Yao, M., Wang, Q., Li, Z,, Yang, L., Huang, P.X., Sun, Y.L., Wang, J., Wang, Y.J., Cui, X.J. (2016). A Systematic Review of Cross-cultural Adaptation of the Oswestry Disability Index. *Spine* (Phila Pa 1976) [Epub ahead of print]

Yardley, L., Sharples, K., Beech, S. (2001) Developing a dynamic model of treatment perceptions. *Journal of Health Psychology.* 6:269.

Zung, W.W. (1965) A self-rating depression scale. *Archives of General Psychiatry* 12:63-70.

Zuse, Konrad (1984) *Der Computer, mein Lebenswerk .* Berlin/Heidelberg: Springer-Verlag.

Glossary of Terms

Area Under the Curve. In this thesis, this is the area under the Receiver Operating Characteristic (ROC) curve.

Bournemouth Questionnaire. This is a short self-report questionnaire containing seven questions. There are separate versions for neck and low back symptoms.

Classical Test Theory. A strategy to measure non-observable constructs by measuring observable characteristics related to the non-observable constructs.

Clinician Based Outcomes. Measures of outcome assessed by a clinician alone.

Clinical Commissioning Groups. These are clinically-led bodies created after the introduction of the Health and Social Care Act, 2012. They are responsible for the planning and commissioning of local health services.

Construct validity. This is the degree to which the scores of a measurement instrument are consistent with hypotheses.

Content validity (including face validity). This is the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured.

Core Outcome Measures in Effectiveness Trials (COMET). An initiative to bring together different parties interested in the development and application of agreed standardised sets of outcome measurement, known as core outcome sets.

Confidence Interval. A range of values around the point estimate that is likely to span the population parameter. They are used to estimate how far away the population mean is likely to be from the sample mean with a degree of certainty.

COnsensus-based **S**tandards for the selection of health **M**easurement **In**struments. The COSMIN initiative aims to improve the selection of health measurement instruments.

Criterion validity. This is the degree to which the scores of a measurement instrument are an adequate reflection of a "gold standard" instrument.

Evidence Based Medicine. Defined by Sackett *et al.,* 2000 as 'the integration of best research evidence with clinical expertise and patient values'.

Effect Size. This is the size of the difference in mean values between two groups, relative to the standard deviation (Barton and Peat, 2014).

ePROM. An electronic version of a Patient Reported Outcome Measure.

Face validity. The degree to which a measurement instrument looks as though it is

an adequate reflection of the construct to be measured.

General Osteopathic Council. The United Kingdom regulator for osteopaths.

Global Perceived Effect. A rating of perceived recovery acquired by asking patients to rate how much their recovery has improved or deteriorated since a predefined time point.

Institute of Osteopathy. The professional association for United Kingdom osteopaths.

Integrated Services Digital Network. A set of communication standards for simultaneous digital transmission of voice, video, data, and other network services over the traditional circuits of the public switched telephone network.

Intraclass Correlation Coefficient. A parameter of agreement or reliability. A value of 1 is perfect, 0.75 is excellent, 0.40 to 0.75 denotes fair to poor performance, and values below 0.40 indicate poor levels of agreement or reliability.

Item Response Theory. A measurement theory that can be applied when the underlying model of measurement is reflective.

Measurement error. This is the systematic and random error in a patient's score. This error cannot be attributed to true changes in a patient being reflected in the construct of interest being measured.

Minimal Clinical Important Change. Former name for Minimal Important Change.

Minimal Clinical Important Difference. Former name for Minimal Important Difference.

Minimal Detectable Change. Former name for Smallest Detectable Change.

Minimal Important Change. The smallest difference in score in the domain of interest which patients perceive as beneficial, and would mandate (in the absence of troublesome side effects and excessive cost) a change in a patient's management.

Minimal Important Difference. The smallest difference between groups which may be considered to be of clinical importance.

National Council for Osteopathic Research. A United Kingdom organisation responsible for promoting a profession wide research culture that is inclusive, robust, credible, has national and increasingly international impact and benefits for osteopathic teaching, learning and patient care.

National Institute of Health and Care Excellence. An organisation providing national guidance and advice to improve health and social care based on the analysis of available research evidence.

Number Needed to Treat. The number of patients that must be treated with an intervention for one extra patient to experience an improvement or benefit using an agreed and established threshold of evaluation.

Numerical Rating Scale. A scale usually ranging from 0-10 marked by a patient to measure a health construct.

Osteopathic Educational Institutions. Training institutions in the United Kingdom which have received Recognised Qualification (RQ) status for the award of BSc or MSc in Osteopathy.

Oswestry Disability Index. A patient reported outcome measure evaluating disability associated with spinal pain.

P- value. A test statistic used to determine the probability of a particular outcome occurring. Probability is expressed as a *p*-value and are set at <0.05 or <0.01 which are generally regarded as being thresholds of statistical significance.

Patient Centred Outcome Measure. A new concept which involves putting patients, and their families and carers, at the heart of deciding which goals are most valuable for individuals with a range of health conditions.

Patient Reported Experience Measure. A measurement instrument for patients to provide direct feedback on their experience of care to drive improvement in services.

Patient Reported Outcome Measure. A form of questionnaire which is completed by patients whose objective is to measure their health status.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses. An evidence-based minimum set of items for reporting in systematic reviews and meta-analyses.

Primary Care Outcomes Questionnaire. The Primary Care Outcomes Questionnaire (PCOQ) is a 24-item questionnaire designed to measure outcomes in primary care. It measures status at a point in time, with change between two points calculated as a difference in scores.

Randomised Controlled Trial. A form of research study design in which participants are randomly allocated to receive a particular intervention. The intervention may be an active substance or activity, am inert product or activity (a placebo), or current standard treatment or substance which acts as a control.

Reliable Change Index. This is the product of dividing the difference between the pre-treatment and post-treatment scores by the standard error of the difference between the two scores.

Research Governance Framework. A framework to complement existing governance arrangements in clinical practice, and financial management in the National Health Service and research community. It is intended to prevent poor performance, adverse incidents, misconduct or fraud and promote public confidence in research.

Receiver Operator Characteristic curve. This is a plot of sensitivity (indicating true positives) against 1-specificity (false positives).

Reliability. This is the degree to which an instrument is free from measurement error. This can be evaluated by the extent to which the scores remain the same when patients whose health status has not changed complete measurement instruments at different points in time. This form of reliability is known as test-retest reliability. Other forms of reliability include inter-rater reliability which can be evaluated to see if a score remains the same when a measurement instrument is used by different people on the same occasion. If an instrument is used by the same person on different occasions and evaluated, this is known as intra-rater reliability.

Responsiveness: The ability of an instrument to detect clinically important change over time within a desired construct.

Roland Morris Disability Questionnaire. A patient reported outcome measure which is back-specific and used to measure levels of patient disability in 12 separate categories. It is available in nine different formats.

Smallest Detectable Change. The level of change that is detectable beyond the measurement error of an instrument. It is inversely proportional to the number of measurement performed by either groups or individuals.

Social Exchange Theory. This is a social psychological and sociological perspective that explains social change and stability as a process of negotiated exchanges between different parties and for different underlying reasons.

Spearman Correlation Coefficient. This is a statistical test used for non-parametric data when evaluating the relationships between different groups being investigated.

Standard Deviation. This is a measure of the dispersion or variability of data. It indicates the difference between a group of values and their mean when all of the data are taken into account.

Standard Error of Measurement. A statistic which describes the measurement error of an instrument *e.g.* a patient reported outcome measure.

Standardised Mean Difference. This is referred to also as the effect size.

Standardised Response Mean. A value calculated by dividing the mean change between groups by the standard deviation of the change scores.

Theory of Planned Behaviour. This is a theory that links beliefs and behaviour by proposing that an individual's attitude toward behaviour, subjective norms, and perceived behavioural control, together shape that individual's behavioural intentions and actions.

Transition Question. This is a single question which asks patients to report whether they have experienced an improvement or deterioration since beginning their treatment. Patients respond using a scale of either 7 or 15 point options.

Type I error. This occurs when a true null hypothesis is rejected, and a false alternative hypothesis is accepted.

Type II error. This occurs when a false null hypothesis is not rejected.

Validity. This is defined by the COSMIN group as the degree to which an instrument measures the construct(s) it purports to measure. Different types of validity should be considered including construct validity, content validity (including face validity), and criterion validity.

Visual Analogue Scale. This a continuous scale which runs from 0-100mm or 0-10cm allowing measurement of a health construct of interest.

Wireless Application Protocol. This is a technical standard for accessing information over a mobile wireless network.

Acronyms

AGREE	Appraisal of Guidelines Research and Evaluation
ALBDS	Aberdeen Low Back Disability Scale
AMED	Allied and Complementary Medicine Database
App	Application
AQP	Any Qualified Provider
ASA	Advertising Standards Authority
ASES	American Shoulder and Elbow Society Scoring System
AUC	Area Under the Curve
BBS	Berg Balance Scale
BOA	British Osteopathic Association
BQ	Bournemouth Questionnaire
CAG	Confidentiality Advisory Group
CAQDAS	Computer Assisted Qualitative Data Analysis Software
CBO	Clinician Based Outcomes
CCG	Clinical Commissioning Groups
CES	Current Employment Statistics
CFI	Comparative Fit Index
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CKS	Cincinnati Knee Scale
CM	Contant-Murley
CNFDS	Copenhagen Neck Functional Disability Scale
CNQ	Core Neck Questionnaire
COMET	Core Outcome Measures in Effectiveness Trials
COPM	Canadian Occupational Performance Measure
COSMIN	**CO**nsensus-based **S**tandards for the selection of health **M**easurement **In**struments
CPD	Continuing Professional Development
CPG	Chronic Pain Grade
CPRD	Clinical Practice Research Datalink

CQC	Care Quality Commission
CSAG	Clinical Standards Advisory Group
CTT	Classical Test Theory
CWOM	Core Whiplash Outcome Measure
DAAG	Data Access Advisory Group
DASH	Disabilities of the Arm, Shoulder and Hand
DCMS	Department for Culture, Media, and Sport
DCV	Discriminant Content Validity
df	Degrees of Freedom
DIF	Differential Item Functioning
DIMDI	Deutschland Institute of Medical Documentation and Information
DO	Diploma in Osteopathy
DRAM	Distress and Risk Assessment Method
EBM	Evidence Based Medicine
ELBPG	European Low Back Pain Guidelines
ELBPWG	European Low Back Pain Working Group
EMBASE	Excerpta Medica Database
EMPRO	Evaluating the Measurement of Patient-Reported Outcomes
ePROM	Electronic Patient Reported Outcome Measure
EQ5D	Euroquol 5 Dimensions questionnaire
ES	Effect Size
FABQ	Fear Avoidance Beliefs Questionnaire
FAQ	Frequently Asked Questions
FCE	Functional Capacity Evaluation
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GCRO	General Council and Register of Osteopaths
GDP	Gross Domestic Product
GIC	Global Impression of Change
GP	General Practitioner
GPE	Global Perceived Effect
GOsC	General Osteopathic Council
GPES	General Practice Extraction Service

GPS	Global Positioning System
GPET	General Practice Extraction Tool Query Service
GRoC	Global Rating of Change
HES	Hospital Episode Statistics
HHS	Harris Hip Score
HOOS	Hip Osteoarthritis Outcome Score
HRA	Health Research Authority
HSCA	Healthcare Supply Chain Association
HSCIC	Health and Social Care Information Centre
IAG	Independent Advisory Group
IASP	International Association for the Study of Pain
ICC	Intraclass Correlation Coefficient
ICHOM	International Consortium for Health Outcomes Measurement
ICO	Office of the Information Commissioner
IIGOP	Independent Information Oversight Group Panel
IMMPACT	Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
iO	Institute of Osteopathy
IQR	Inter Quartile Range
IRT	Item Response Theory
ISDN	Integrated Services Digital Network
ISOQoL	International Society for Quality of Life Research
IT	Information Technology
LAQ	Long Arc Quadriceps
LBOS	Low Back Outcome Score
LEFS	Lower Extremity Functional Scale
KAKPS	Kujala Anterior Knee Pain Scale
KMO	Kaiser-Meyer-Olnik coefficient
KOOS	Knees injury and Osteoarthritis Outcome Score
KPMG	Klynveld Peat Main Goerdele
LBPRS	Low Back Pain Rating Scale
LILACS	Literatura Latino Americana em Ciências da Saúde
MCIC	Minimal Clinical Important Change

MCID	Minimal Clinical Important Difference
MCS	Minimum Change Score
MDC	Minimum Detectable Change
MHRA	Medicines and Healthcare products Regulatory Agency
MIC	Minimal Important Change
MID	Minimal Important Difference
MMICS	Multinational Musculoskeletal Inception Cohort Study
MOS	Medical Outcomes Study
MOT	Medical Outcomes Trust
MRC	Medical Research Council
MRO	Member of the Register of Osteopaths
MskHQ	Musculoskeletal Health Questionnaire
MSPQ	Modified Somatic Pain Questionnaire
MVAS	Million Visual Analogue Scale
NASS LSO	North American Spine Surgeons Lumbar Spine Outcome Assessment Instrument
NatCen	National Centre for Social Research
NDI	Neck Disability Index
NHS	National Health Service
NCOR	National Council for Osteopathic Research
NICE	National Institute of Health and Care Excellence
NIH	National Institute for Health
NNT	Number Needed to Treat
NPDS	Neck Pain and Disability Scale
NPNPQ	Northwick Park Neck Pain Questionnaire
NRS	Numerical Rating Scale
ODI	Oswestry Disability Index
OECD	Organisation for Economic Cooperation and Development
OEI	Osteopathic Educational Institutions
OFCOM	Office of Communications
OFT	Office of Fair Trading
OMERACT	Outcome Measures in Rheumatoid Arthritis Clinical Trials
OMT	Osteopathic Manipulative Treatment

ONS	Office for National Statistics
PARIHS	Promoting Action on Research Implementation in Health Services
PASE	Physical Activity Scale for the Elderly
PCOM	Patient Centred Outcome Measure
PCOQ	Primary Care Outcomes Questionnaire
PCT	Primary Care Trust
PDA	Personal Digital Assistant
PE	Patient Experience
PEDro	Physiotherapy Evidence Database
PIS	Participant Information Sheet
PLC	Pain Locus of Control
POEM	Patient Outcome and Experience Measure
PORT	Patient Outcomes Research Teams
PPT	Pain Pressure Threshold
PREM	Patient Reported Experience Measure
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient Reported Outcome Measure
PROMIS	Patient Reported Outcome Measure Information System
PS	Patient Satisfaction
PSFS	Patient Specific Functional Scale
QBPDS	Quebec Back Pain Disability Scale
QMREC	Queen Mary University of London Research Ethics Committee
QOF	Quality and Outcomes Framework
QUOROM	Quality Of Reporting Of Meta-analyses
RCC	Royal College of Chiropractors
RCGP	Royal College of General Practitioners
RCI	Reliable Change Index
RCT	Randomised Controlled Trial
RGF	Research Governance Framework
RoM	Range of Motion
RMDQ	Roland Morris Disability Questionnaire
ROC	Receiver Operator Characteristic
RODQ	Revised Oswestry Disability Questionnaire

SaaS	Software as a Service
SCC	Spearman Correlation Coefficient
SCS	Standardised Change Score
SD	Standard Deviation
SDC	Smallest Detectable Change
SDQ	Shoulder Disability Questionnaire
SEM	Standard Error of Measurement
SET	Social Exchange Theory
SF-36	Short Form 36
SIGLE	System for Information on Grey Literature in Europe
SIP	Sickness Impact Profile
SPADI	Shoulder Pain and Disability Index
SPSS	Statistical Package for the Social Sciences
SRM	Standardised Response Mean
SSL	Secure Sockets Layer
StaRI	Standards for Reporting Phase IV Implementation Studies
TLI	Tucker Lewis Index
TPB	Theory of Planned Behaviour
TQ	Transition Question
TUG	Timed Up and Go
TRT	Test-retest
UK	United Kingdom
UKBEAM	United Kingdom Back Pain, Exercise and Manipulation
VAS	Visual Analogue Scale
VRS	Verbal Rating Scale
WAP	Wireless Application Protocol
WDI	Waddell Disability Index
WHO	World Health Organisation
WHOQoL	World Health Organisation Quality of Life
WOMAC	Western Ontario and McMasters University Arthritis Index
WOOS	Western Ontario Arthritis of the Shoulder Index
ZON	Zong Onderzoek Nederland

Statement of originality

Required statement of originality for inclusion in research degree theses

I, [Carol Fawkes], confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: C.A.Fawkes

Date: 06-11-2016