BENCHMARKING OUTCOMES FOR PSYCHOLOGICAL TREATMENTS OF CHRONIC PAIN

Grania Fenton

Submitted in accordance with the requirements for the degree of
Doctor of Clinical Psychology (D. Clin. Psychol.)

The University of Leeds
Academic Unit of Psychiatry and Behavioural Sciences
School of Medicine

June 2010

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to Sharon Barlow, Jeanette Barnes, Amber Baxter, Dee Burrows, Claire Copland, Reuben Crossley, Ben Davies, Conor Fenton, Dominic Gage, Nadine Hobro, Daniel Lockley, Matthew Mawer, Ken Obbard, Nicola Smart, Dave Walsh, Maggie Whittaker, and all of the clinicians working in busy Pain Management Programmes who took time out from their busy schedules to contribute to this research.

I would also like to thank the authors of clinical trials who responded to my queries, and, of course, Professor Stephen Morley, for his unwavering enthusiastic supervision of this research.

ABSTRACT

In an attempt to bridge the widely acknowledged gap between research and clinical practice, this thesis examined the feasibility of benchmarking outcomes for published psychological treatments of chronic pain for application within routine clinical settings. Benchmarking outcomes is relatively common for psychological treatments in the mental health field, but in spite of the prevalence of chronic pain and its known impact on many areas of functioning, the chronic pain literature has previously only considered the generic application of benchmarks for developing services and considering standards for waiting times.

Four studies of mixed methodological approaches were conducted. The first aimed to ascertain the extent of similarities between published psychological treatments of chronic pain and treatments delivered in routine clinical settings. This was to ensure that the application of benchmarks from the published literature to routine clinical settings would be meaningful. The second study examined whether the published literature was likely to facilitate the development of benchmarks, and the third sought clarification in terms of outcome domains within which useful benchmarks could be generated. The final study was a meta-analysis of data extracted from the published literature within specified outcome domains.

The results suggested that it would be meaningful to apply benchmarks produced from the published literature to routine clinical settings, and that the literature would facilitate the development of benchmarks within several outcome domains. The meta-analysis led to the generation of four benchmarks. These were in the outcome domains of pain experience and physical functioning when compared with waiting list controls, and coping and cognitive appraisal and emotional functioning when compared with active controls. The impact of the design of each study and properties inherent within the literature on the benchmarks generated and their application within routine clinical settings was then considered, prior to suggestions for future research and clinical applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACT: Acceptance and Commitment Therapy

BDI: Beck Depression Inventory

CBT: Cognitive Behavioural Therapy

CFS: Chronic Fatigue Syndrome

CI: Confidence Intervals

CSQ: Coping Strategies Questionnaire

HADS: Hospital Anxiety and Depression Scale

ITT: Intention to Treat sample

PICT: Psychologist in Clinical Training

PMP: Pain Management Programme

PSEQ: Pain Self-Efficacy Questionnaire

RCT: Randomised Controlled Trial

SD: Standard Deviation

SIG: Special Interest Group

SM: Professor Stephen Morley

SPSS: Statistical Package for the Social Sciences

TAU: Treatment as usual

WLC: Waiting list control

INTRODUCTION

The clinical psychology profession is based on the 'scientist-practitioner' model; developed in a drive to overcome the conflict between clinical practice and research (Barkham & Mellor-Clark, 2003) by ensuring the ability of practitioners to function competently as both clinicians and researchers (Barker, Pistrang & Elliott, 2002). The accreditation of clinical psychology training courses aims to facilitate the development of such competencies, in insisting that in addition to clinical skills, psychologists in clinical training (PICTs) are taught how to digest research and apply findings to clinical practice, and are supervised in conducting advanced applied research within a doctoral level thesis (British Psychological Society, 2007). This thesis aimed to both demonstrate the author's ability to function as a researcher, and attempt to bridge a gap between research and clinical practice by considering the published research data into psychological treatments of chronic pain and its potential application to clinical practice. The overall aim was to use the published literature to benchmark outcomes for psychological treatments of chronic pain which could be applied within clinical settings, and a number of steps were taken in order to achieve this. These included determining whether the application of research data to clinical settings would be meaningful and if the published literature would facilitate such application; considering the outcome domains within which benchmarks could be generated; and finally, ascertaining what the benchmarks suggested about the impact of psychological treatments of chronic pain and considering how they could be used within clinical practice.

Research and Clinical Practice

The National Institute for Health and Clinical Excellence considers the best available evidence in its clinical guidance, and because Randomised Controlled Trials (RCTs) and systematic reviews and meta-analyses of RCTs are considered to provide evidence of the highest quality (Greenhalgh, 2006), they tend to have the greatest impact on clinical guidance. RCTs produce 'efficacy' data about treatments, and the application of such data to routine clinical settings is referred to as 'evidence-based practice'. Although this paradigm is often assumed to be sufficient to support practice in routine clinical settings, the controlled conditions reported in efficacy studies mean that data is not necessarily externally valid to clinical settings. This has resulted in a 'gap' between research and clinical

practice, and explains why it has been argued that the move towards evidence-based practice has essentially re-emphasised the research-practice division within the clinical psychology profession (Kazdin, 2008). In contrast to evidence-based practice, 'practice-based evidence' comprises data about therapeutic effectiveness obtained from clinical settings, and as such is regarded as externally valid. Whilst the methodological quality of effectiveness studies is not considered to be comparable to that of RCTs, effectiveness findings are at least representative of the populations to which they will be applied (Seligman, 1995; Shadish, Matt, Navarro & Phillips, 2000).

Efficacy and effectiveness studies have their own strengths and weaknesses, but it has long been recognised that research into the efficacy as well as effectiveness of treatments is essential to ensure that the evidence-base directing practice does not suppress innovation (NHS Executive, 1996). Green and Glasgow (2006) have recently acknowledged that that 'if we want more evidence-based practice, we need more practice-based evidence' (p. 126),and there is therefore a need to acknowledge both RCT and clinical setting data, and to attempt to overcome the inherent difficulties in translating findings across these settings (Kazdin, 2008).

Bridging the Gap between Research and Clinical Practice: Benchmarking

One way in which efficacy data can be translated to clinical settings is through the use of 'benchmarking' (Barkham, Mellor-Clark, Connell & Cahill, 2006; Minami, Wampold, Serlin, Kircher & Brown, 2007; Weersing & Weisz, 2002). Benchmarking allows professionals within clinical settings to consider research evidence (Goodheart, et al., 2006), and reflect on its application to clinical practice within a comprehensible format (Minami, Serlin, Wampold, Kircher & Brown, 2008; Wampold, Goodheart & Levant, 2007). Clinical governance requires the effectiveness of routine clinical practice to be measured (Department of Health, 1999), and benchmarking allows this in terms of the extent to which clinical outcomes are comparable to research outcomes (Minami, et al., 2008; Wade, Treat & Stuart, 1998); a comparison which can also facilitate an understanding of 'what works for whom' (Roth & Fonagy, 1996).

Benchmarking Psychological Treatments

Whether or not benchmarks aggregate efficacy data to compare with aggregated effectiveness data or utilise single studies to compare outcomes, benchmarking can

enhance the knowledge base for psychological treatments within a process considered to be a 'virtuous circle' (Barkham, Mellor-Clark, Connell & Cahill, 2006). This process not only informs clinical practice, but is also congruent with the clinical psychology profession's scientist-practitioner role (Barkham & Mellor-Clark, 2003). Additionally, the ability of benchmarking to accommodate for the limitations of effectiveness designs was first acknowledged three decades ago (T. Cook & Campbell, 1979), and more recently it has also been credited with helping to overcome the limitations of efficacy data (Wade et al.,1998).

There are a number of psychological therapies and diagnostic categories across which benchmarking has been applied. The seminal study of Wade et al. (1998) focused on panic disorder, and the magnitude of treatment effects was estimated within an effectiveness study which was comparable with two efficacy studies. The measure of effectiveness within this study was 'relative' in that it was determined by the proportion of participants meeting pre-set criteria, relative to the proportion of participants demonstrating similar improvements within the efficacy studies. Wade et al. (1998) found that the magnitude of effectiveness was on a par with those found in the efficacy studies, with the vast majority of participants (87%) describing themselves as 'panic free' post-treatment. Similar levels of improvement were seen in terms of anticipatory and general anxiety and levels of depression.

Wade et al. (1998) can be considered to have set the precedent for subsequent benchmarking studies, which have in the main been conducted within the depression literature in examining eclectic services with adolescent (Weersing & Weisz, 2002) and adult populations (Barkham, Rees, Leach, Shapiro, Hardy & Lucock, 2005; Barkham, Stiles, Connell, Twigg, Leach, Lucock, et al., 2008); and adults receiving a variety of standardised therapies (Minami, et al., 2007). Benchmarking has more commonly been applied within the adult mental health field, but has also been used within health psychology to determine the relative effectiveness of a pilot programme of Cognitive Behavioural Therapy (CBT) for clients with Chronic Fatigue Syndrome (CFS) to RCT outcomes (Scheeres, Wensing, Knoop & Bleijenberg, 2008) and to produce quality guidelines for standards of care for people with epilepsy (Pugh, et al., 2007).

Chronic Pain

Pain is subjective, and has been defined by Merskey and Bogduk (1994) as 'an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage' (p. 210). Chronic pain is technically pain which has lasted for six months or longer (APA, 2004), although most literature considers pain that does not resolve, is unresponsive to treatment and lasts for longer than three months to be an indicator of chronicity (Tunks, Crook & Weir, 2008).

The Impact of Chronic Pain

The estimated prevalence of chronic pain in adults is between 12% and 30% (Breivik, Collett, Ventafridda, Cohen & Gallacher, 2006). It is associated with significant distress and disability in all domains of life (Turk, Swanson & Tunks, 2008); contributing to somatic complaints and problems with social and emotional well-being and abilities to manage daily activities (Breivik, et al., 2006). Chronic pain has negative implications for employment and work related roles (Boersma & Linton, 2006; Breivik, et al., 2006; Morley & Eccleston, 2008), and is associated with fatigue, poor sleep, reliance on support systems, loss of social and familial roles, anger and frustration (Morley & Eccleston, 2008) and depression (Boersma & Linton, 2006; Breivik, et al., 2006; Morley & Eccleston, 2008; Tunks, Weir & Crook, 2008). People with chronic pain also frequently report sexual dysfunction associated with depression, poor coping skills and inadequate social support, with indirect effects of medication on impairment (Kwan, Roberts & Swalm, 2005).

Pathology does not predict an individual's perceptions of and responses to pain but evidence suggests that cognitive, emotional, behavioural and social factors impact on both pain experience and response to treatment (Adams, Poole & Richardson, 2006; Turk, 1999; Turk & Okifuji, 2002); with employment, coping styles (Tunks, Weir, et al., 2008), marital satisfaction, spousal responses to pain (Cano, Gillis, Heinz, Geisser & Foran, 2004), feelings of control over pain, resilience (Karoly & Ruehlman, 2006), and fear-avoidance and pain beliefs (Boersma & Linton, 2006) all being identified as moderators of outcome.

Treatments of Chronic Pain

Given the detrimental effects of chronic pain on multiple areas of an individual's life and the many moderators of outcome, a biopsychosocial model is essential in

understanding the experience and ensuring that effective treatments are provided (Adams, et al., 2006; Philips & Rachman, 1996; Tunks, Crook, et al., 2008; Turk & Okifuji, 2002; Turk, Swanson, et al., 2008). This is acknowledged within recommendations that multi-disciplinary treatments are provided for people with chronic pain (British Pain Society, 2007; Tunks, Weir, et al., 2008; Turk & Okifuji, 2002).

Treatments for chronic pain include psychodynamically informed insight-oriented approaches, motivational interviewing, biofeedback, guided imagery and hypnosis (Turk, Swanson, et al., 2008); and treatments such as behavioural therapy, biofeedback and relaxation are all associated with improvements to pain experience, affect, coping and social role functioning when compared with control groups (Morley, Eccleston & Williams, 1999). Evidence also suggests that mindfulness meditation is associated with reduced physical and psychological pathology, with many changes being maintained at follow-up (Kabat-Zinn, Lipworth & Burney, 1985). Mindfulness meditation also accounts for significant proportions of variance in terms of pain, physical, social, emotional and cognitive functioning, overall disability and medication use (McCracken, Gauntlett-Gilbert & Vowles, 2007). Similarly, acceptance-based approaches have been associated with reduced healthcare use and improvements to levels of social, physical and emotional functioning (McCracken, Vowles & Eccleston, 2005).

More than fifty published RCTs have examined the efficacy of psychological treatments of chronic pain, with CBT demonstrating the greatest empirical support on a significant number of pain related domains including pain, affect and disability (Eccleston, Williams & Morley, 2009; Morley, et al., 1999) and coping strategies (Morley, et al., 1999). There is also emerging evidence for the effectiveness of CBT within routine clinical settings (Morley, Williams & Hussain, 2008).

Benchmarking Psychological Treatments of Chronic Pain

The meaningful application of benchmarking has been demonstrated across a number of treatment settings. However, this does not challenge the intuitive argument that benchmarking is of greatest utility when treatments delivered are standardised and available in clinical and research settings, such as Beck's model of cognitive therapy for depression (Beck, Ward, Mendelson, Mock & Erbaugh, 1961), and the same outcome measures are employed within both settings. Benchmarking may also be of greatest use

when treated populations are similar enough to one another to ensure that research and clinical comparisons are meaningful. Psychological treatments of chronic pain, in clinical or research settings, do not lend themselves to such comparisons however. Rather, the literature reports myriad psychological treatments, methods of delivery, levels of therapist training and experience, treatment populations and outcome measures. It has been acknowledged that the comparison of data generated within RCT and clinical settings is not ideal given that the settings are unlikely to be identical, but that until an alternative becomes available, benchmarking efficacy data is the best available option (Minami, et al., 2009).

Prior to considering benchmarking for psychological treatments of chronic pain, the existing benchmarking literature and ways in which it has overcome heterogeneity obstacles warrants consideration, in order that the feasibility of producing benchmarks within an imperfect context can be explored. It demonstrates that studies have varied, for example, in terms of whether they chose to focus on single or multiple psychological treatments. Wade et al. (1998) focused on CBT but aggregated data from trials regardless of whether delivery had been individual or in groups, and Weersing and Weisz (2002) examined the same broad psychological treatment as deemed to be the 'research standard of care' regardless of eclectic modes of delivery. Other studies used data from any treatments that were evidence-based (Pugh, et al., 2007), or those which had been identified as the primary treatment within a recent literature review (Scheeres, et al., 2008). Outcomes from different therapies have also been aggregated (Barkham, Rees, et al., 2005; Barkham, et al., 2008; Minami, et al., 2007), and this mode of benchmarking is permitted given the reported equivalent effects of different psychological treatments that are based on psychological principles and delivered by qualified therapists (Wampold, et al., 1997).

Another factor worthy of consideration is the observed heterogeneity of levels of experience and professional backgrounds of therapists delivering treatments within the chronic pain literature (Eccleston, et al., 2009; Morley, et al., 1999). This also requires consideration in terms of both the way in which benchmarking is conducted and the external validity of the findings.

With reference to treatment populations, the chronic pain literature tends to categorise clients by presentation to services rather than diagnosis (Morley, 2008; Morley

& Eccleston, 2008) as both anecdotal and research findings suggest that diagnosis does not account for a great deal of variance in treatment outcome (Van Tulder, Assendelft, Koes & Bouter, 1997). Most benchmarking studies have considered populations with a homogeneous diagnosis, however, with the exception of Barkham et al. (2008) who included more heterogeneous groups, and such methods are likely to possess greater external validity and be more valuable within the chronic pain literature.

The benchmarking literature has also demonstrated variations in terms of methods employed. For example, Scheeres et al. (2008) reported the proportion of patients within their study who demonstrated clinically significant levels of improvement. This reflects a recent trend towards more sophisticated methods of demonstrating change in terms of whether post-treatment scores were within normative ranges; methods also employed by Barkham et al. (2008) and Wade et al. (1998). As Scheeres et al. (2008) were unable to access similar data within efficacy studies, however, they also calculated the magnitude of the treatment effect within all studies using effect sizes and confidence intervals (CIs); therefore enabling them to make relative comparisons between each treatment. Barkham, Rees et al. (2005) and Barkham et al. (2008) made similar relative treatment comparisons, but chose to transform scores between homogeneous measures used in clinical settings and compare these with a standard measure used in research settings; methods that are unfortunately irrelevant to the chronic pain literature given the heterogeneity of measures employed (Eccleston, et al., 2009; Morley, et al., 1999). Weersing and Weisz (2002) additionally compared their outcomes with treatments described in the efficacy literature, and, like Minami et al. (2007), reported the relative effectiveness of their treatment to control groups. Such relative methods are of great utility, as if comparisons are not made with a control group, the magnitude of the treatment effect is likely to be overestimated as there is no allowance for spontaneous remission.

Whilst matters of treatments, therapist and population factors and outcome measures require considerable forethought, it is reasonable to assume that chronic pain is another field of health psychology within which benchmarking may be feasible. It has previously been considered in terms of designing services (Donovan, Evers, Jacobs & Mandleblatt, 1999) and specifying waiting times within this field (Lynch, et al., 2008) and a recent consultation exercise aims to generate general best practice benchmarks for the treatment of pain (Department of Health, 2009). However, benchmarking has yet to be

considered as one way of bridging the gap between research and clinical practice concerning outcomes for psychological treatments of chronic pain. Bridging this gap is important for a number of reasons, not least to ensure that treatments provided in clinical settings are evidence based and facilitate realistic treatment expectations. Research suggests that whilst patient reported levels of satisfaction with treatments provided for chronic pain, whether medical, physical or psychological, are generally good (George & Hirsh, 2005; Hirsh, et al., 2005; McCracken, Evon & Karapas, 2002), levels of satisfaction with the effects of treatment vary between 40% (Breivik, et al., 2006) and 55% (Hirsh, et al., 2005). This may be related to recent findings that patients often have unrealistic expectations of the effects of psychological treatments (Thorne & Morley, 2009). Though it is recognised that expectations often change as treatments progress, unrealistic expectations are somewhat understandable given that neither clinicians nor researchers are currently able to convey realistic outcomes. Expectations are known to impact on psychological outcomes and pain experience however (Kirsch, 1985, 1999; Price, et al., 1999), and chronic pain is admittedly a relatively common condition which can have a detrimental impact on many aspects of functioning. There is therefore a clear argument for at least attempting to build a bridge between research and clinical practice; aiming to both provide effective treatments and generate realistic outcomes as a result.

Research Questions

The literature reviewed led to the development of a number of research questions related to the issue of benchmarking outcomes for psychological treatments of chronic pain:

1. **Is it meaningful to use the efficacy literature to develop benchmarks which can be applied to routine clinical settings?**

   This was explored within study 1, which aimed to determine whether routine clinical practice is evidence-based, and subsequently, whether it would be meaningful to compare it with published efficacy data from RCTs (see pages 20-30).

2. **Does the published literature contain data which facilitates the development of benchmarks?**

This question was examined within study 2, which aimed to determine whether the aggregation of data from published RCTs is likely to produce benchmarks (see pages 31-44).

3. **Are there particular outcome domains within which it is useful to produce benchmarks for use in clinical settings?**

   Study 3 aimed to answer this question in generating a consensus opinion of clinicians working in routine clinical practice as to useful domains to benchmark (see pages 45-56).

4. **What do the benchmarks suggest about the impact of psychological treatments of chronic pain within particular outcome domains?**

   This question was examined within study 4: a meta-analysis of the data identified in study 2 within the domains identified within study 3 (pages 57-81).

As there were a number of research questions to be explored within four separate studies, a mixed methodological approach was taken. Each study has been described sequentially in terms of its aims, methods, results and a study-specific discussion prior to a final overall discussion. It is worth noting, however, that as several of the procedures were conducted concurrently, some studies make reference to data obtained in later studies. With this in mind, every effort has been taken to ensure that such references are both logical and necessary.

STUDY 1: IS IT MEANINGFUL TO USE THE EFFICACY LITERATURE TO DEVELOP BENCHMARKS WHICH CAN BE APPLIED TO ROUTINE CLINICAL SETTINGS?

AIMS

This study aimed to determine whether there are enough similarities with the published literature to ensure that benchmarks developed in efficacy contexts could be meaningfully applied to clinical settings.

METHOD

Design

This study employed a mixed quantitative and qualitative design which focused on the staff, patients, treatments and outcome measures within UK based Pain Management Programmes (PMPs).

Ethical Clearance

This study did not require ethical approval, as confirmed in Appendix 1.

Participants

Participants were members of staff working in PMPs across the UK.

Measures

A survey was designed in order to determine whether routine clinical practice is evidence-based (see Appendix 2). The survey questions were based on a form for extracting data from RCTs for chronic pain. This form had been previously approved by the Cochrane Collaboration and used within a recent systematic review of the chronic pain literature (Eccleston, et al., 2009).

Procedure

Participants were invited to participate by several methods. Professor Stephen Morley (SM) introduced the study and invited participation following his plenary session at the British Pain Society's Annual Scientific Meeting in April 2009. Additionally, an email was sent to approximately 200 people; seventy-six of whom were members of the British Pain Society's

PMP Special Interest Group (SIG) (British Pain Society, 2009) and the rest of whom had attended the most recent PMP SIG conference and volunteered their contact details.

The survey was developed and deployed using the 'Bristol Online Survey' method; a widely accessible internet-based survey method used by over one-hundred higher education institutes and public and private sector companies (University of Bristol, 2009). An electronic link to the survey was sent to potential participants, who were invited to respond within a reasonable time frame. Responses were summarised and to ensure that the anonymity of participants was maintained, a summary of the preliminary results was forwarded to the original email list, whether or not they had chosen to participate (see Appendix 3). The results were then compared with preliminary data extracted from the published literature identified within study 2, with specific reference to the psychological treatment within each RCT that had been classified as primary for the purposes of this research.

RESULTS

Thirty-six separate responses were received in total, all of which were from people involved with services in England. All responses were included within the preliminary results sent to participants, but the responses of three participants were excluded from the final analysis of data because two participants worked within services that were not classified as PMPs and one participant chose to withhold the name of the service within which they worked and so its PMP status could not be determined. Some participants chose to not respond to particular questions, but all proportions were calculated in relation to the overall maximum number of available responses, i.e. thirty-three.

Whilst the survey was designed in consideration of an approved data extraction form, the various ways in which PMPs collect and publicise data means that an exact comparison between PMP and RCT data is not possible. Instead, a qualitative summary of the PMP data is provided in order to supplement the RCT data reported in Table 1 and the available comparisons in Table 2. As such, this was a descriptive study.

Participant Characteristics

More than two-thirds of participants (n=22) stated that they were responding as individuals, with the remaining eleven stating that their responses were on behalf of their team. Twenty-four participants (73%) were members of the British Pain Society, and the majority (67%; n=22) were physiotherapists (n=13) or clinical psychologists (n=9). The

remaining eleven participants (33%) classified themselves as belonging to the medical, nursing or occupational therapy professions.

Team Characteristics

Thirty-one of the participants (94%) classed themselves as belonging to an 'inter-disciplinary' team where multi-disciplinary professionals worked together on a routine basis. The results indicated that these teams frequently comprised administrative staff, clinical psychologists, physiotherapists and members of the medical and nursing professions. Several respondents also stated that assistant psychologists and PICTs, CBT therapists, counselling and health psychologists, graduate patients, pharmacists and occupational therapists made contributions to the team. Such multi-disciplinary teams were also commonly reported within the RCT literature.

In terms of professional development, 67% of participants (n=22) stated that an induction programme was routine for new staff and that systematic clinical supervision was routinely provided. Whilst only 36% (n=12) stated that therapeutic training was routinely provided for clinicians, 85% (n=28) agreed that staff were encouraged to allocate time for continuing professional development on a regular basis. Several participants suggested that the manner in which continuing professional development, clinical supervision, inter-disciplinary working and induction periods were encouraged and provided was variable.

Patient Characteristics

Participants' estimates of the number of patients seen per annum varied, reflecting the varied capacities and clinical foci of PMPs. Most estimated that fewer than 100 patients were treated per annum (52%; n=17), 36% (n=12) estimated between 100 and 200, and 12% (n=4) estimated more than 200. Similar data were not extractable from the RCTs, but all participants estimated that the average age of patients was between 40 and 50, and as Table 1 demonstrates, this was not dissimilar to the average age amongst participants entered into the RCTs (mean 51.5; SD 11.7). Services did not appear to have imposed upper age limits, but only 6% of respondents (n=2) implied that referrals of patients under the age of 18 would be accepted. Both PMP and RCT settings seemed to comprise a larger proportion of female than male patients, as 85% of participants (n=28) estimated that at least 60% of patients were female; similar to the 72% of RCTs shown in Table 1 (n=26) comprising samples where at least 60% of patients were female.

Table 1: Included studies and preliminary sample and outcome data

Study	Mean age (SD)	% female	Primary diagnosis or site of pain	Follow-up data collected
Altmaier 1992	39.9 (8.9)	27	CLBP	•
Astin 2003	47.7 (10)	99	MSK (Fib)	•
Basler 1997	49.3 (9.7)	59	CLBP	•
Becker 2000	56.5 (15)	64	Mixed	•
Bliokas 2007	45.2 (9.6)	56	Mixed	○
Carson 2005	51.1 (nk)	60	CLBP	○
Cook 1998	77.6 (nk)	62	Mixed	•
Ersek 2003	82 (nk)	87	Mixed	○
Ersek 2008	81.9 (nk)	85	Mixed	•
Evers 2002	53.7 (nk)	71	RA	•
Flor 1993	42.4 (9.7)	64	Mixed	•
Greco 2004	47.3 (10.4)	93	SLE	•
Haldorsen 1998	43 (10.6)	64	Mixed	○
Jensen 2001	43.3 (10.4)	56	Mixed	•
Johansson 1998	43.5 (7.6)	78	MSK	•
Keefe 1990	64 (11.5)	72	OA	•
Keefe 1996	62.6 (10.1)	61	OA	•
Kraaimaat 1995	57 (12.7)	68	RA	•
Linton 2008	47.5 (8.6)	55	Spi	○
Mishra 2000	35.7 (9.9)	82	TMD	○
Moore 1985	49.3 (13.2)	98	Mixed	•
Morone 2008	74.9 (6.1)	57	CLBP	○
Newton-John 1995	45.4 (11.6)	61	CLBP	•
Nicassio 1997	53.1 (nk)	89	MSK (Fib)	•
Puder 1988	52.7 (14.4)	71	Mixed	•
Radojevic 1992	54.4 (nk)	76	RA	•
Thieme 2003	47.6 (nk)	100	MSK (Fib)	•
Turner 1988	46 (nk)	59	CLBP	•
Turner 1990	44 (nk)	48	CLBP	•
Turner 1993	42 (nk)	54	CLBP	•
Turner 2006	37.1 (11.3)	81	TMD	•
Vlaeyen 1995	42.2 (nk)	75	CLBP	•

Study	Mean age (SD)	% female	Primary diagnosis or site of pain	Follow-up data collected
Vlaeyen 1996	44 (9.4)	88	MSK (Fib)	○
Williams 1996	50 (nk)	56	Mixed	•
Woods 2008	46.5 (11.9)	66	CLBP	•
Zautra 2008	52.4 (nk)	68	RA	○

Note • = measured, ○ = not measured ; % female = percentage of patients in each sample who were female, calculated from published information therefore sometimes at start of treatment, sometimes at end; CLPB = Chronic Low Back Pain; Fib = Fibromyalgia; MSK = Musculoskeletal; nk = not known; OA = Osteoarthritis; RA = Rheumatoid Arthritis; SLE = Systemic Lupus Erythematosus; Spi = spinal; TMD = Temperomandibular.

The preliminary RCT data extracted within study 2 also allowed for some specific diagnostic comparisons to be made. For example, 21% of participants (n=7) made specific reference to the fact that it was common for patients with multiple sites of pain and mixed diagnoses to be seen within their PMP (with service inclusion criteria exerting the main influence). The RCT literature also made reference to the multiple sites of pain experienced by many patients and the influence of service inclusion criteria, with 31% of studies (n=11) involving patients with mixed diagnoses.

Twenty-one of the participants (64%) estimated that at least 61% of their patients commonly had chronic low back pain, and nearly a third of the included RCTs (28%; n=10) involved patients with this primary diagnosis. Sixteen participants (48%) estimated that at least 70% of their patients had musculoskeletal pain, and 14% of the included RCTs (n=5) might therefore apply to them. Both the RCT and PMP data implied that osteoarthritis was a less common diagnosis amongst patients treated, as nearly half of all PMP respondents estimated that around 11% of their patients had this diagnosis, and only two of the included studies (6%) involved patients with this primary diagnosis. Amongst the remaining included RCTs, rheumatoid arthritis was the primary diagnosis within four studies (11%), temperomandibular joint syndrome the focus of two (6%) and systemic lupus erythematosus of one (3%). The remaining study involved patients with primary spinal pain.

Participants suggested that fewer patients were likely to present with diagnoses such as complex regional pain syndrome, facial pain, headache, inflammatory arthropathy such as rheumatoid arthritis or lupus, temperomandibular joint syndrome, vascular, post-operative or visceral pain.

Treatment Characteristics

Participants were asked to describe the components of the main psychological treatment offered within their PMP, and so for comparative purposes, data regarding a psychological treatment deemed to be primary was extracted from the RCTs included in study 2. Table 2 demonstrates the proportion of characteristics within PMP treatments and primary psychological treatments characteristics from included RCTs.

Table 2: Comparison of treatment and outcome characteristics between PMPs and RCTs

Characteristic		**PMP data (%)**	**RCT data (%)**
Primary Treatment			
	Out-patient	88	86
	Total duration at least 10 hours	91	94
	CBT [a]	39	61
	CBT plus other components	18	11
	Mixed/other/unknown primary treatment [b]	43	28
	Group based	61	89
	Individual based	0	11
	Involvement of significant others	58	16
	Mixed methods of delivery [c]	27	28
Components of treatment			
	Educational	97	100
	Relaxation	88	80
	Cognitive restructuring	85	72
	Behavioural management	73	72
	Goal-setting	97	69
	Coping-skills	97	69
	Exercise	94	47
	Problem-solving	88	50
	Mindfulness	67	8
	Biofeedback	30	8
Outcome measures			
	Administered pre-treatment	94	100
	Post-treatment	94	100
	Follow-up	85	75

Note
[a] CBT = Cognitive Behavioural Therapy. [b] Unknown treatment refers to the response of one participant within the PMP survey.
[c] mixed methods of delivery refers to treatments that, for example, may involve the delivery of some components in a group, other components on an individual basis etc.

PMPs and RCTs are clearly comparable with regard to the provision of treatments within out-patient settings using mixed methods of delivery of at least ten hours duration. The mean number of hours treatment was higher in PMPs (51) than RCTs (28.9), although the large standard deviation (SD) in RCTs (34.1) reflects the variability in treatment duration which ranged from six to 160 hours. As participants gave approximate estimates of treatment duration it was not possible to calculate the SD from the mean of PMP treatment duration, but

they estimated that the main psychological treatments provided were between three and over one-hundred hours in duration. This indicates that both research and clinical settings involve treatments of varying duration.

A greater proportion of the primary psychological treatments in RCTs than PMPs were classified as CBT or CBT in addition to other components (72% and 57% respectively). Whilst the proportion of PMPs using such treatments was significant, several participants described a recent move to more 'third wave' cognitive behavioural treatments such as Acceptance and Commitment Therapy (ACT) and mindfulness based cognitive therapy, particularly when pain was not related to activity (n=1) or where many patients did not have English as a first language (n=1).

Similar proportions of non-CBT primary treatments were reported within the PMP and RCT data, and a large proportion of all primary treatments in both settings were provided within groups, although 11% of the RCTs (n=4) described individual psychological treatments. PMPs were more likely than RCTs to involve significant others within the main psychological treatment, and seventeen of the nineteen participants who stated this implied the involvement of significant others within specific treatment sessions rather than as integral to the treatment as a whole.

A large proportion of PMPs and RCTs cited education, relaxation, cognitive restructuring, behavioural management, goal setting and coping skills as components of the main psychological treatment, although the latter two were more common within PMPs. Additionally, exercise, problem solving, mindfulness and biofeedback were more commonly components of the main PMP treatments than the main RCT treatments.

Within PMPs, 48% of participants (n=16) reported that manuals were used within the main psychological treatment, and 44% (n=7) implied that they guided all aspects of treatment. RCTs were more likely to refer to manuals (61%; n=22), although on closer inspection, 91% of these studies (n=20) had adapted treatments that were described within published papers rather than following a specified manualised protocol. Similarly, all manuals used within PMPs had been developed in-house and specifically tailored in line with the treatment aims.

Whilst it was not possible to determine the proportion of RCTs who offered follow-up appointments to patients, 91% of PMP participants (n=30) stated that their programme offered follow-ups, with 39% (n=13) stating that they routinely followed up patients who dropped out of treatment; usually by telephone or in writing.

Outcome Characteristics

Similar proportions of RCTs and PMPs administered outcome measures both before and after treatment (94% of PMPs and 100% of RCTs), with a slightly greater proportion of PMPs than RCTs administering follow-up measures (85% and 75% respectively).

Participants reported that outcome measures within the domains of coping, mood, disability and health and social care use were most likely to be administered within PMPs, and all participant-reported measures of coping, mood, disability and pain experience were standardised. There was some consistency within the coping domain within which the Coping Strategies Questionnaire (CSQ) of Rosensteil and Keefe (1983) and the Pain Self-Efficacy Questionnaire (PSEQ) of Bandura (1997) were commonly used. The Hospital Anxiety and Depression Scale (HADS) designed by Zigmond and Snaith (1983) was the most commonly employed measure of mood. Within the domain of health and social care use, employment status, medication use and number of GP and other medical service visits were commonly recorded, and there was less consistency amongst measures of disability.

DISCUSSION

This study sought to determine the extent of the similarities between the published literature of psychological treatments of chronic pain and routine clinical practice in terms of team, patient, primary psychological treatment and outcome characteristics. Its ultimate aim was to ascertain whether routine clinical practice is evidence-based, and therefore if it would be meaningful for clinical settings to apply benchmarks developed from the efficacy literature.

The results imply that, at least amongst the PMPs sampled, clinical practice is similar enough to the RCT literature to be considered 'evidence-based' for a number of reasons. Firstly, the multi-disciplinary nature of teams within both settings is promising, and reflects an initial comparability. Secondly, the range of estimated diagnoses amongst patients within PMPs was similar to those reported within RCTs. Even though the estimated representation of patients with particular diagnoses in PMPs was not equivalent to that reported in RCTs, anecdotal reports and research suggesting that diagnosis does not exert a major impact on outcome (Van Tulder, et al., 1997) imply that this does not provide a cause for concern. Patient groups were also estimated to comprise people with a similar average age and have similar levels of female predominance. Thirdly, treatment characteristics were commonly comparable; particularly when the main psychological models employed, treatment settings

and ranges of duration were considered. Given that the primary psychological treatments in both PMPs and RCTs were seen to be CBT plus components of other treatments, the significant overlap of many treatment components is perhaps not surprising. Finally, the timing of the administration of outcome measures was similar in both settings. Whilst the outcome domains were considered further in studies two, three and four and so were not discussed within this study, all in all, the results of this study give credence to the argument that it would be meaningful to develop benchmarks from efficacy data that could be applied within routine clinical settings.

Although it was not possible to compare the exact team and patient characteristics (e.g. the number of patients treated annually within services described in RCTs) and so the level of resources in both settings could not be determined, these results suggest that there is a 'good enough' level of similarities amongst PMPs and the published literature to ensure that any benchmarks produced would be meaningful. A number of factors require further consideration, however, including the discrepancies noted amongst treatment components. With this in mind, it is noteworthy that one of the consequences of benchmarking by aggregating data from many studies is that the external validity of the results may be compromised. However, unless, following the example of Wade et al. (1998), PMPs develop services based on treatments provided within the published literature, benchmarking will always be carried out within sufficiently comparable clinical and research contexts rather than exactly comparable contexts (Minami, et al., 2009). The methods of determining treatment components within the RCT data may also have contributed to such discrepancies and similarities, as data were only examined for psychological treatments that, for the purposes of this research, were deemed to be primary. The decision making process regarding the classification of primary treatments is described within study 2, and such a process was necessary to manage the volume of data generated by the included studies and ensure that it led to meaningful benchmarks. However, whilst it facilitated a straightforward comparison between the two sets of data, an unfortunate consequence of such a focus is a lack of understanding of all treatments offered within both settings. Another confounding factor may have been the RCT authors' reporting of treatment components, as a lack of explicit mention of any component meant that it was deemed to be lacking, whether or not this was in fact the case. The scope of this research did not permit a more thorough comparative description of treatments in both settings, as it aimed to ascertain feasibility, but future research may benefit

from describing components of all treatments provided within both PMPs and RCTs in order to gain a more thorough appreciation of the similarities between treatments offered in both settings, and therefore the extent to which research findings have external validity.

Another potential methodological limitation of this study was the fact that current PMP activity (i.e. at the time of data collection in 2009) was compared with historical RCT data, as studies had been published between 1985 and 2008. The description of treatment specific components may have overcome this to an extent, but until the volume of published data is sufficient enough to warrant only comparing recent PMP activity with recently published data, such a limitation is difficult to overcome.

In further consideration of the methods employed, the potential for bias amongst the self-selected sample of participants within this descriptive study needs to be considered. For example, although anonymity was assured, the 33% of participants (n=11) who stated that they were responding on behalf of their team may have felt the need to portray their PMP in a particular manner, and as such the results should be considered as an indicator of PMP practice rather than conclusive evidence. The fact that 67% of respondents (n=22) considered themselves to be responding as individuals within their PMP also may have influenced the results and needs consideration. The decision to give participants the option to respond either as an individual or on behalf of their PMP was a deliberate attempt to encourage participation even if responders were unable to discuss the survey responses with colleagues, but a 'collective' PMP response may also have been very different. Another potential source of bias may have been profession-related, as most participants were physiotherapists or clinical psychologists and may have provided significantly different responses to colleagues of other professions. In future, a larger sample from a wider range of professionals would provide a more representative set of results and hopefully reflect the multi-disciplinary teams described within PMPs and RCTs. This may involve seeking more precise data, using more than descriptive methods, within a service evaluation context however, and the potential impact on participant recruitment would require some forethought. In the meantime, this study should be considered to have provided an indication of PMP activity and how appropriate it is to apply RCT data to such contexts, based on estimates from a small sample of participants within approximately 41% of the 87 PMPs in England (Waring & Booth, 2006).

Finally, whilst in so far as it was possible, every effort was taken to determine whether it would be meaningful to apply the efficacy literature to PMPs, without analysing the specific

aims and detailed content of the interventions provided within both settings the actual comparability can only ever be assumed. It is known that RCTs tended to adapt treatments described within published papers, and PMPs are more likely to develop in-house manuals, but the precise content of all treatments remains unknown. In future, a greater transparency of the components of interventions tested within efficacy trials would mean that they had greater clinical utility and facilitate more 'fine-grained' comparisons to be made, and perhaps the publication of accessible (e.g. internet based) manuals would facilitate comparisons in this manner (Thorn, Cross & Walker, 2007). In response to Green and Glasgow's (2006) argument that 'if we want more evidence-based practice, we need more practice based evidence' (p. 126) though, perhaps PMPs also need to be more transparent regarding their treatment aims and interventions; and this may also be more feasible within a service evaluation context, where the publication of data is considered.

STUDY 2: DOES THE PUBLISHED LITERATURE CONTAIN DATA WHICH FACILITATES THE DEVELOPMENT OF BENCHMARKS?

AIMS

Study 1 demonstrated that in principle, it would be meaningful to use the efficacy literature to develop benchmarks which can be applied to routine clinical settings. This study aimed to determine whether or not the aggregation of data from the published literature was likely to produce benchmarks.

BACKGROUND LITERATURE

In attempting to achieve this study's aim, it was fundamental that high quality data were identified within the published literature. Several methods of review were therefore considered. The systematic review methods of the 'Cochrane Review' are considered to produce the highest standard of evidence-based research as they apply pre-defined and rigorous methods for searching the literature (The Cochrane Collaboration, 2010). Such systematic methods involve a thorough search of published primary literature and are considered scientific as they aim to identify as much of the relevant literature as possible. Such methods also control for bias and random error due to specified inclusion criteria and the statistical methods employed. Pre-defining literature search methods within Cochrane reviews also allows readers to make judgements about the quality of the review and ensure that methods are both open to criticism (DeCoster, 2004) and replicable (Centre for Reviews and Dissemination, 2001, 2009; Davies & Crombie, 2001; Glass, 1976; Lipsey & Wilson, 2001; Quintana & Minami, 2006; White, 1994).

Alternatively, narrative reviews may be conducted. These tend to be as comprehensive as systematic reviews but subjective interpretation of the data are likely, whereas within systematic reviews, the reader should be able to draw the same conclusions as the author (Centre for Reviews and Dissemination, 2001, 2009). Although both types of review involve aggregating the literature and weighting the included studies in order that studies employing larger samples contribute a greater proportion to the overall results, systematic reviews rely on more sophisticated systematic procedures for doing so (Lipsey & Wilson, 2001). Narrative methods also tend to solely rely on statistical significance measures, despite the recognised unreliability of such measures within research (Quintana & Minami, 2006). They are also more

appropriate when research questions are broad rather than focused as they facilitate a qualitative description of study results (Centre for Reviews and Dissemination, 2001; D. Cook, Mulrow & Haynes, 1997; Quintana & Minami, 2006). In contrast, benchmarking necessitates the quantitative analyses of results using meta-analysis; and only a systematic literature review can facilitate such methods (Centre for Reviews and Dissemination, 2001, 2009).

METHOD

Design

A consideration of the available methods of review and the chronic pain literature led to a decision to extract data from studies identified within the 2009 Cochrane review of Eccleston et al.

Ethical Clearance

As identifying and manipulating the literature to conduct benchmarks involved the use of secondary data within the public domain, this study did not require the approval of an ethics committee.

Measures

A modified version of the data extraction form used in the Eccleston et al. (2009) review was developed. This enabled data about study design, participants, diagnoses, treatments and outcome measures to be extracted (see Appendix 4). A data extraction code book was also developed so that data regarding treatment arms and outcome measures could be coded appropriately (Appendix 5).

Procedure

Selection Criteria

The original selection criteria of Eccleston et al. (2009) had been designed to facilitate the statistical comparison of post-treatment change between active psychological treatments and other active treatments or controls from waiting list control (WLC) or treatment as usual (TAU) samples. This research, however, aimed to identify data which could be used to produce benchmarks of pre-post change. A priori selection criteria regarding participants, psychological treatments and outcomes were therefore modified slightly in order that studies providing

appropriate pre, post and, if applicable, follow-up data be included (the methods of meta-analysis which this data facilitated, and their justification, are reported within study 4). In practice, this meant that studies were deemed appropriate for selection if they contained at least one treatment with a specified psychological component which was led, or supervised by, a professional with psychological training, and had also employed a suitable control group. In order to ensure that benchmarks were as meaningful to PMPs as possible, 'control' groups were classified as groups that would be viable within PMPs, and so were either from a WLC or specified TAU sample, or another sample deemed to be reasonable within PMPs. Only data from studies with more than ten independent participants within each treatment or control group were included.

Data Collection

All studies identified within the 2009 Cochrane review were examined for eligibility for inclusion. Following the initial selection, only those facilitating the extraction of appropriate data were included. The types of outcome measures typically employed within the chronic pain literature and associated suitable methods of meta-analysis had been considered, and outcome measures scored on continuous scales were deemed appropriate.

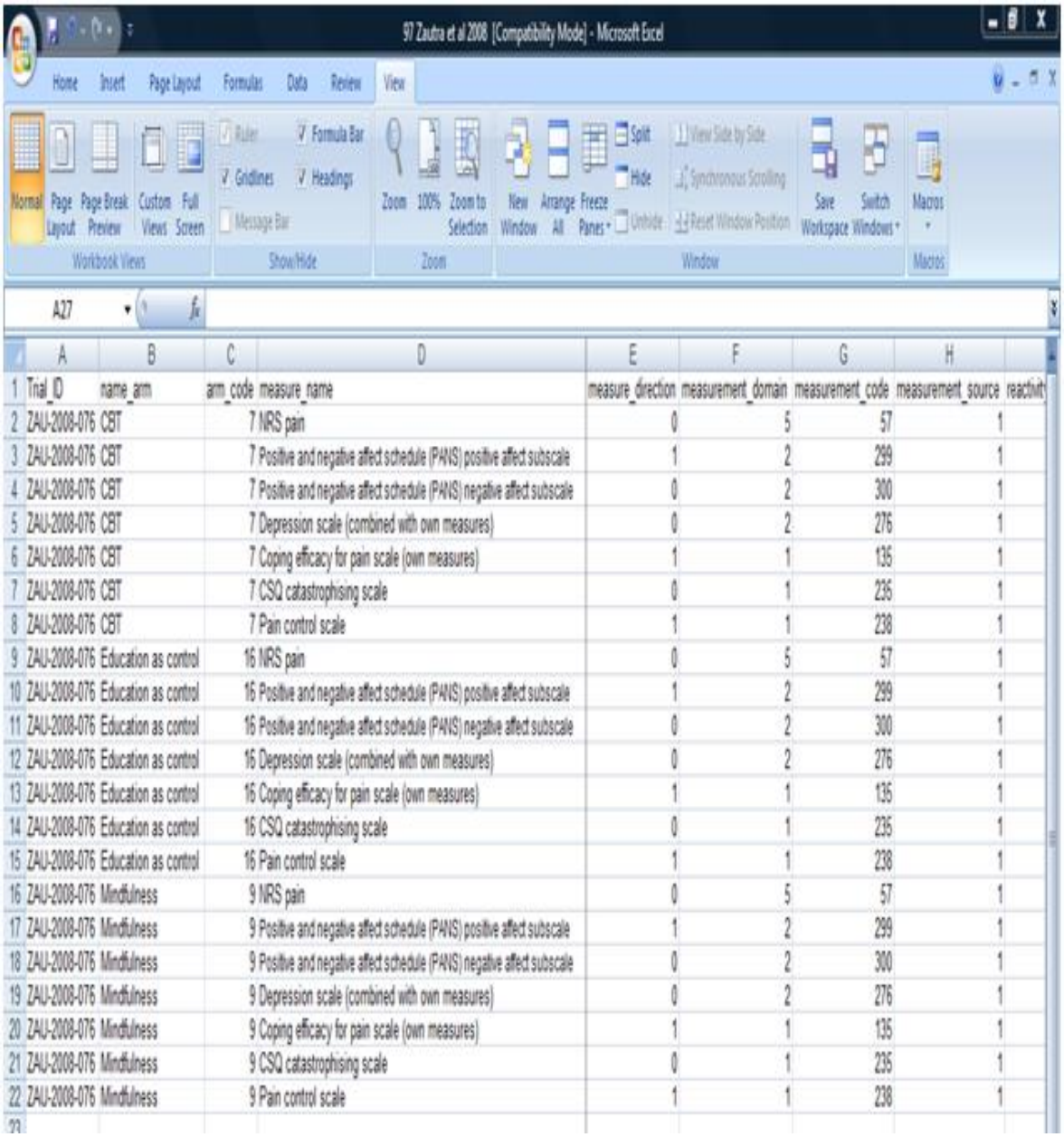
The first stage of data collection involved identifying studies reporting raw mean pre and post-treatment scores for treatment and control groups, and pre-treatment and control SDs, or data enabling the calculation of pre-treatment SDs (e.g. pre-treatment standard errors of the mean). Raw scores are known to produce more conservative estimates of effect size than pre-post mean change scores (Morris & De Shon, 2002), but if such data were not available, studies were still deemed appropriate for inclusion if they reported pre-post mean change scores and pre-treatment SDs (as such data facilitated the calculation of effect sizes). Where suitable data were not published, primary authors were contacted. If they did not respond within a reasonable time frame, the study was excluded from the analysis; based on the approaches used in previous research (Eccleston et al., 2009; Morley et al., 1999).

As the methods by which data identified from systematic literature reviews are stored, organised and analysed are considered to be the most important factors in meta-analytic data management (Cooper & Hedges, 1994), guidance recommends that data identified is managed within software which has the ability to code data (Centre for Reviews and Dissemination, 2009). However, as the coding of data evolved as data were extracted and unfamiliar issues

became more overt, a less complex software package which facilitated the entry of data from more than two treatment arms and multiple outcome measures was necessary for initially storing data. Microsoft Excel was considered to provide the most appropriate storage medium, and data from each study was extracted into an individual Microsoft Excel file, and pooled if necessary. The evolving code book was used to develop 'value labels' within an SPSS (Statistical Package for the Social Sciences) version 15 file. To minimise the risk of data extraction errors, each Excel file was separately transferred to SPSS and visually scanned for errors and omissions. Examples of the Excel and SPSS files can be seen in Figures 1 and 2. In order to ensure reliability, SM checked a proportion of data for errors, and each SPSS file was then transferred into a single SPSS file where frequency analyses and other checks were run to identify missing data and that which had been entered erroneously.

The classification of primary psychological treatments and control arms was also discussed at this stage, as preliminary data extraction demonstrated the potential volume of data that was available within studies given their frequent reporting of more than one treatment and control arm. The subsequent need to be selective for pragmatic reasons and to ensure that benchmarks could be as of much utility to PMPs as possible was therefore identified. All suitable data were extracted from the included studies, and then supervision was used to discuss how to classify primary psychological treatment and control arms in studies with more than one of each arm. The predominance of treatments based on CBT principles and with the involvement of significant others identified in study 1 meant that they were classed as primary if they were available; with CBT with no involvement of significant others otherwise classified as primary. A decision was also made to prioritise out-patient over in-patient treatments if necessary due to their predominance within PMPs. Amongst studies with more than one control arm, WLCs were selected as primary if available, as most PMPs have access to waiting lists and therefore could feasibly identify comparable data. If studies had not employed a WLC, a decision was made as to whether a specified TAU or another active control (e.g. education only) would be feasible within PMPs, and if studies reported more than one active control then the least intensive one was classified as primary.

Figure 1: Screen shot of Microsoft Excel data extraction file



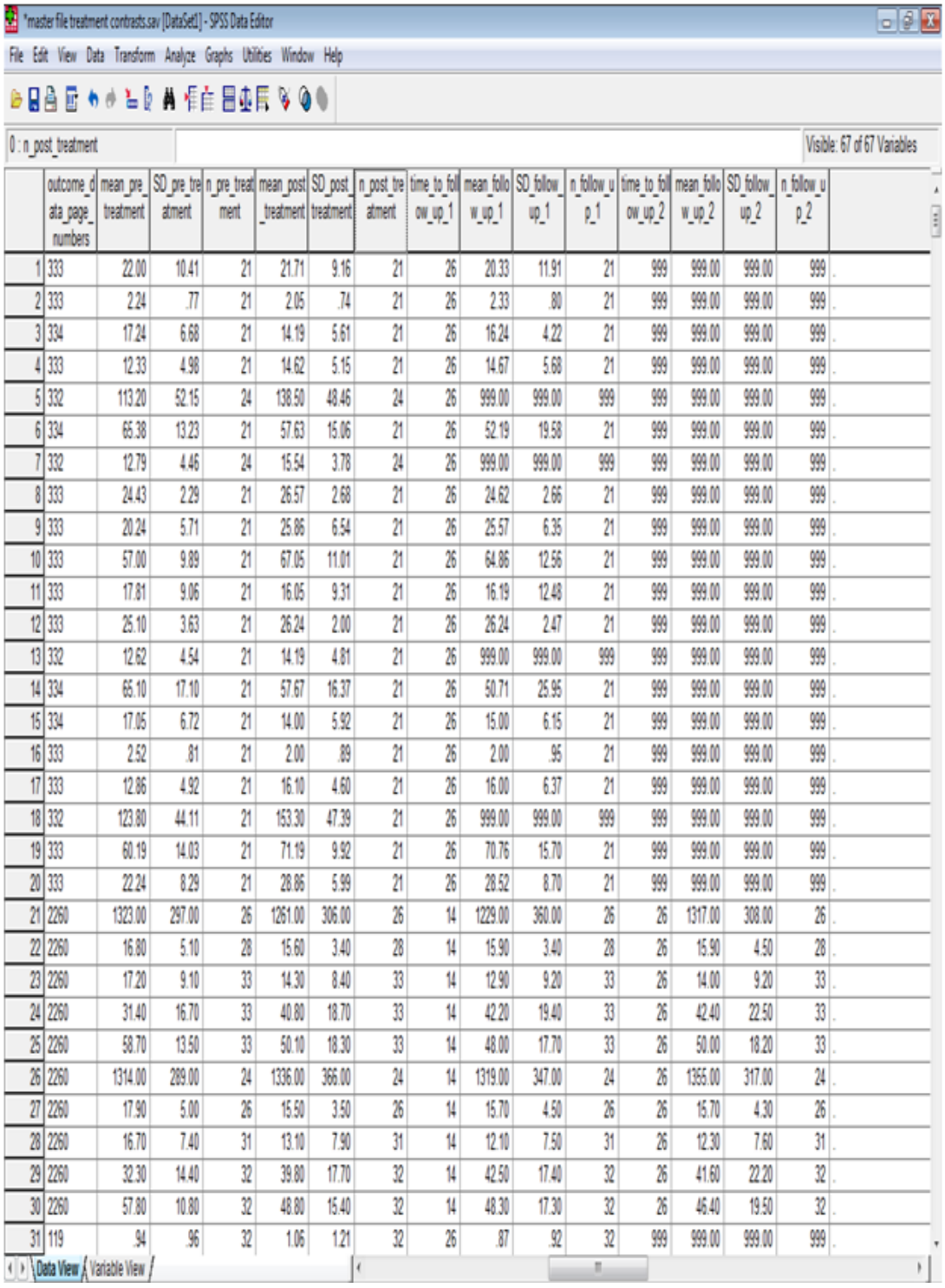
97 Zautra et al 2008 [Compatibility Mode] - Microsoft Excel

Home Insert Page Layout Formulas Data Review View

A27

	A	B	C	D	E	F	G	H	
1	Trial_ID	name_arm	arm_code	measure_name	measure_direction	measurement_domain	measurement_code	measurement_source	reactivity
2	ZAU-2008-076	CBT	7	NRS pain	0	5	57	1	
3	ZAU-2008-076	CBT	7	Positive and negative affect schedule (PANS) positive affect subscale	1	2	299	1	
4	ZAU-2008-076	CBT	7	Positive and negative affect schedule (PANS) negative affect subscale	0	2	300	1	
5	ZAU-2008-076	CBT	7	Depression scale (combined with own measures)	0	2	276	1	
6	ZAU-2008-076	CBT	7	Coping efficacy for pain scale (own measures)	1	1	135	1	
7	ZAU-2008-076	CBT	7	CSQ catastrophising scale	0	1	235	1	
8	ZAU-2008-076	CBT	7	Pain control scale	1	1	238	1	
9	ZAU-2008-076	Education as control	16	NRS pain	0	5	57	1	
10	ZAU-2008-076	Education as control	16	Positive and negative affect schedule (PANS) positive affect subscale	1	2	299	1	
11	ZAU-2008-076	Education as control	16	Positive and negative affect schedule (PANS) negative affect subscale	0	2	300	1	
12	ZAU-2008-076	Education as control	16	Depression scale (combined with own measures)	0	2	276	1	
13	ZAU-2008-076	Education as control	16	Coping efficacy for pain scale (own measures)	1	1	135	1	
14	ZAU-2008-076	Education as control	16	CSQ catastrophising scale	0	1	235	1	
15	ZAU-2008-076	Education as control	16	Pain control scale	1	1	238	1	
16	ZAU-2008-076	Mindfulness	9	NRS pain	0	5	57	1	
17	ZAU-2008-076	Mindfulness	9	Positive and negative affect schedule (PANS) positive affect subscale	1	2	299	1	
18	ZAU-2008-076	Mindfulness	9	Positive and negative affect schedule (PANS) negative affect subscale	0	2	300	1	
19	ZAU-2008-076	Mindfulness	9	Depression scale (combined with own measures)	0	2	276	1	
20	ZAU-2008-076	Mindfulness	9	Coping efficacy for pain scale (own measures)	1	1	135	1	
21	ZAU-2008-076	Mindfulness	9	CSQ catastrophising scale	0	1	235	1	
22	ZAU-2008-076	Mindfulness	9	Pain control scale	1	1	238	1	
23									

Figure 2: Screen shot of SPSS data extraction file



*master file treatment contrasts.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

0 : n_post_treatment

Visible: 67 of 67 Variables

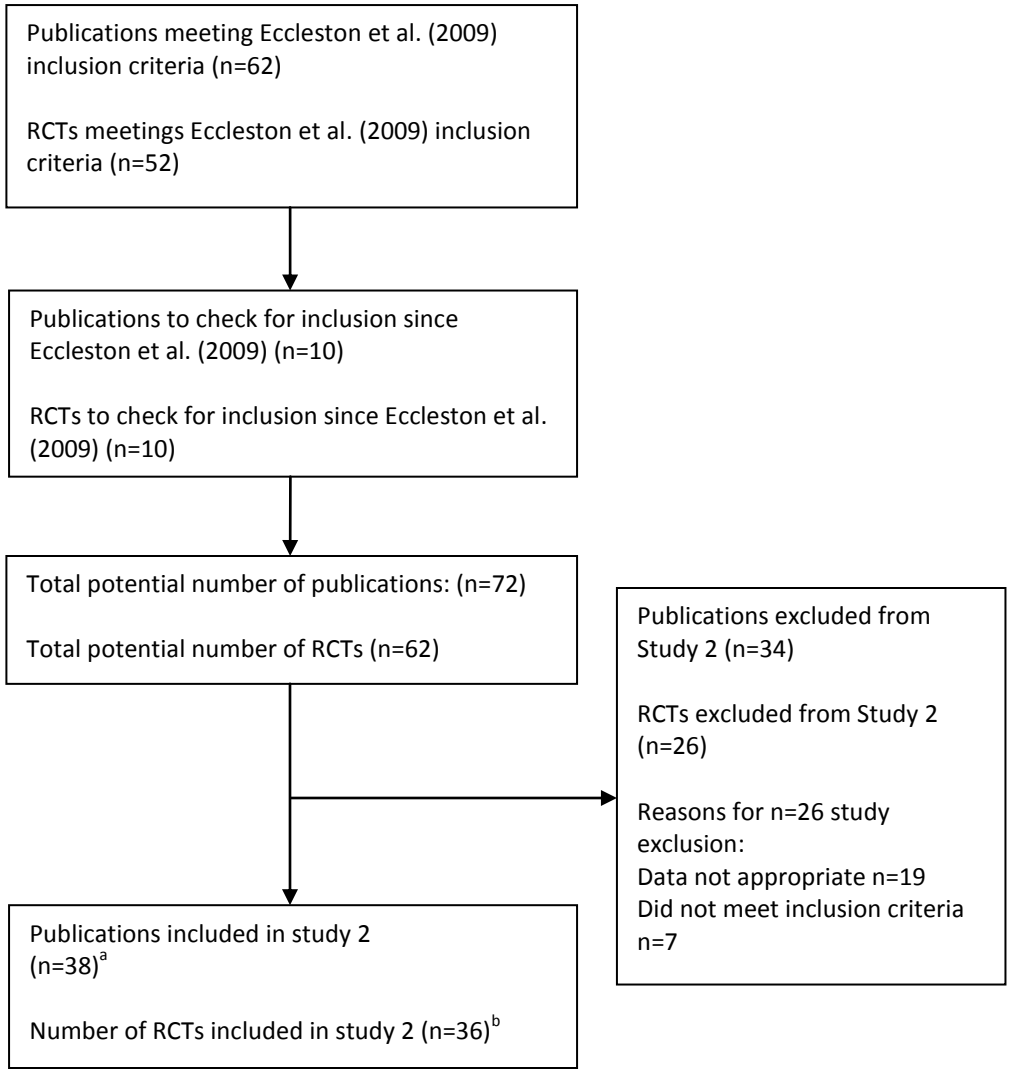
	outcome_data_page_numbers	mean_pre_treatment	SD_pre_treatment	n_pre_treatment	mean_post_treatment	SD_post_treatment	n_post_treatment	time_to_follow_up_1	mean_follow_up_1	SD_follow_up_1	n_follow_up_1	time_to_follow_up_2	mean_follow_up_2	SD_follow_up_2	n_follow_up_2
1	333	22.00	10.41	21	21.71	9.16	21	26	20.33	11.91	21	999	999.00	999.00	999
2	333	2.24	.77	21	2.05	.74	21	26	2.33	.80	21	999	999.00	999.00	999
3	334	17.24	6.68	21	14.19	5.61	21	26	16.24	4.22	21	999	999.00	999.00	999
4	333	12.33	4.98	21	14.62	5.15	21	26	14.67	5.68	21	999	999.00	999.00	999
5	332	113.20	52.15	24	138.50	48.46	24	26	999.00	999.00	999	999	999.00	999.00	999
6	334	65.38	13.23	21	57.63	15.06	21	26	52.19	19.58	21	999	999.00	999.00	999
7	332	12.79	4.46	24	15.54	3.78	24	26	999.00	999.00	999	999	999.00	999.00	999
8	333	24.43	2.29	21	26.57	2.68	21	26	24.62	2.66	21	999	999.00	999.00	999
9	333	20.24	5.71	21	25.86	6.54	21	26	25.57	6.35	21	999	999.00	999.00	999
10	333	57.00	9.89	21	67.05	11.01	21	26	64.86	12.56	21	999	999.00	999.00	999
11	333	17.81	9.06	21	16.05	9.31	21	26	16.19	12.48	21	999	999.00	999.00	999
12	333	25.10	3.63	21	26.24	2.00	21	26	26.24	2.47	21	999	999.00	999.00	999
13	332	12.62	4.54	21	14.19	4.81	21	26	999.00	999.00	999	999	999.00	999.00	999
14	334	65.10	17.10	21	57.67	16.37	21	26	50.71	25.95	21	999	999.00	999.00	999
15	334	17.05	6.72	21	14.00	5.92	21	26	15.00	6.15	21	999	999.00	999.00	999
16	333	2.52	.81	21	2.00	.89	21	26	2.00	.95	21	999	999.00	999.00	999
17	333	12.86	4.92	21	16.10	4.60	21	26	16.00	6.37	21	999	999.00	999.00	999
18	332	123.80	44.11	21	153.30	47.39	21	26	999.00	999.00	999	999	999.00	999.00	999
19	333	60.19	14.03	21	71.19	9.92	21	26	70.76	15.70	21	999	999.00	999.00	999
20	333	22.24	8.29	21	28.86	5.99	21	26	28.52	8.70	21	999	999.00	999.00	999
21	2260	1323.00	297.00	26	1261.00	306.00	26	14	1229.00	360.00	26	26	1317.00	308.00	26
22	2260	16.80	5.10	28	15.60	3.40	28	14	15.90	3.40	28	26	15.90	4.50	28
23	2260	17.20	9.10	33	14.30	8.40	33	14	12.90	9.20	33	26	14.00	9.20	33
24	2260	31.40	16.70	33	40.80	18.70	33	14	42.20	19.40	33	26	42.40	22.50	33
25	2260	58.70	13.50	33	50.10	18.30	33	14	48.00	17.70	33	26	50.00	18.20	33
26	2260	1314.00	289.00	24	1336.00	366.00	24	14	1319.00	347.00	24	26	1355.00	317.00	24
27	2260	17.90	5.00	26	15.50	3.50	26	14	15.70	4.50	26	26	15.70	4.30	26
28	2260	16.70	7.40	31	13.10	7.90	31	14	12.10	7.50	31	26	12.30	7.60	31
29	2260	32.30	14.40	32	39.80	17.70	32	14	42.50	17.40	32	26	41.60	22.20	32
30	2260	57.80	10.80	32	48.80	15.40	32	14	48.30	17.30	32	26	46.40	19.50	32
31	119	.94	.96	32	1.06	1.21	32	26	.87	.92	32	999	999.00	999.00	999

Data View | Variable View

RESULTS

The Eccleston et al. (2009) review had included 62 papers referring to 52 RCTs. This study additionally examined papers whose inclusion was yet to be confirmed when the review was published. In total, 72 publications reporting 62 RCTs were examined for inclusion, and 38 publications reporting 36 separate studies were deemed eligible. Figure 3 displays the selection of studies for inclusion, and information about included and excluded studies is displayed in Tables 3 and 4 respectively.

Figure 3: Flow chart demonstrating the selection of Randomised Controlled Trials (RCTs)



[a] of which 5 had not yet been classified within the 2009 Cochrane review
[b] of which 6 had not yet been classified within the 2009 Cochrane review

Included Studies

Table 3: Characteristics of included studies

Primary author and year of publication	N*	Primary**	Psychological Treatment Active treatment 2	Psychological Treatment Active treatment 3	Duration (hours)	Control WLC	Control Active
Altmaier 1992	45	CBT[a]	o	o	nk	o	TAU
Astin 2003	128	Mind	o	o	20	o	Edu
Basler 1997	94	CBT	o	o	30	o	TAU
Becker 2000	166	CBT	o	o	nk	●	Med
Bliokas 2007	143	CBT	CBT-ex	o	66.5	●	o
Carson 2005	43	LKM	o	o	12	o	TAU
Cook 1998	28	CBT	o	o	10/12[b]	o	Edu
Ersek 2003	45	CBT	o	o	10.5	o	Edu
Ersek 2008	256	CBT[a]	o	o	10.5	o	Edu
Evers 2002	64	CBT	o	o	11	o	TAU
Flor 1993	120	CBT	BFB	o	8	o	TAU
Greco 2004	92	CBT-ex	o	o	nk	o	Sym, TAU
Haldorsen 1998	469	CBT	o	o	120	o	TAU
Jensen 2001	214	CBT	CBT-ex	BT-ex	54, 80, 140[c]	o	TAU
Johansson 1998	42	CBT[a]	o	o	160	●	o
Keefe 1990	99	CBT	o	o	15	o	Edu, TAU
Keefe 1996	88	CBT-ex	CBT	o	20	o	Edu
Kraaimaat 1995	77	CBT	o	o	20	●	TAU
Linton 2008	34	BT	o	o	nk	●	o
Mishra 2000	94	CBT	CBT-ex	BFB	18, 24[d]	o	TAU
Moore 1985	43	CBT-ex	CBT	o	32	●	o
Morone 2008	37	Mind	o	o	12	●	o

Primary author and year of publication	N*	Primary**	Psychological Treatment Active treatment 2	Active treatment 3	Duration (hours)	Control WLC	Control Active
Newton-John 1995	44	CBT	BFB	○	16	●	○
Nicassio 1997	86	BT	○	○	15	○	Edu
Puder 1988	69	CBT	○	○	20	●	○
Radojevic 1992	59	CBT-ex	CBT	○	6	○	Edu, TAU
Thieme 2003	61	BT[a]	○	○	75	○	TAU
Turner 1988	81	CBT	BT	○	16	●	○
Turner 1990	96	BT-ex	BT	○	16	●	PT
Turner 1993	102	CT	CTR	○	12	●	R
Turner 2006	158	CBT	○	○	nk	○	Edu
Vlaeyen 1995	71	CT	BT	BFB	16.5	●	○
Vlaeyen 1996	125	CT	○	○	18, 24[e]	●	Edu
Williams 1996	121	CBT	CBT[a]	○	28, 144[f]	●	○
Woods 2008	44	CBT	BT	○	12	●	○
Zautra 2008	143	CBT	Mind	○	16	○	Edu

Note. The primary authors are for the primary references describing these studies: the Keefe 1990 and Keefe 1996 studies were described in one additional paper each.

*N is pre-intervention and does not necessarily match with the number of participants providing data for analysis.** the classification of treatment and control arms are for the purposes of analysis and do not necessarily reflect the primary authors' classifications.

[a] treatments provided within in-patient settings
[b] the CBT treatment was of 10 hours duration, the education was of 12 hours duration
[c] the CBT treatment was of 54 hours duration, the BT-ex was of 80 hours duration and the CBT-ex was of 140 hours duration
[d] the CBT and BFB conditions were of 18 hours duration, the CBT-ex was of 24 hours duration.
[e] the CT was of 18 hours duration, the Edu of 24 hours duration.
[f] the CBT was of 28 hours duration, the in-patient CBT was of 144 hours duration

• = measured; ◦= not measured/not applicable; BFB = Biofeedback; BT = Behavioural Therapy; BT-ex = Behavioural Therapy plus extra components (Jensen 2001 and Turner 1990 = additional exercise components); CBT = Cognitive Behavioural Therapy; CBT-ex = CBT plus extra components (Bliokas 2007 = graded exposure; Greco 2004 and Mishra 2000 = Biofeedback; Jensen 2001 = behavioural and physical components; Keefe 1996, Moore 1985 & Radojevic 1992= CBT with the involvement of significant others) ; CT = Cognitive Therapy; CTR = Cognitive Therapy and Relaxation; Edu: Education; LKM = Loving Kindness Meditation; Med = medical; Mind = Mindfulness; nk = not known; PT = Physical Therapy; R = Relaxation; Sym = Symptom Monitoring; WLC = Waiting List Control.

Patients

As data regarding the patient diagnoses, ages and proportions of gender from the included studies was reported in Table 1 (study 1) it was not repeated within Table 3. It is worth remembering, however, that study 1 demonstrated the mean age of patients within included studies to have been 51.5 (SD 11.7) with a mean age range of between 35.7 and 81.9. It had also shown that all included studies reported the percentage of patients who were female, which ranged from 29 to 100% (mean 67.7; SD 16), and that the primary diagnostic labels reported within studies were: mixed (31%); chronic low back pain (28%); musculoskeletal (14%); rheumatoid arthritis (11%); osteoarthritis (5%), temperomandibular (5%) systemic lupus erythematosus (3%) and spinal (3%).

Additionally, Table 3 demonstrates that a total of 3681 patients were entered into the thirty-six included studies, with a mean number of entered patients per study of 102.25 (SD 81.13). At the end of treatments, 3160 patients completed outcome measures within these studies, giving a crude estimate of a drop-out rate of 14%.

Treatments

Sixteen of the included studies (44%) reported one active psychological treatment and one control arm, twenty-one (58%) had presented data from only one active psychological treatment, twelve (33%) from two active psychological treatments and three (8%) from three active psychological treatments. The classification of a psychological treatment as primary was necessary for the 42% of included studies (n=15) reporting more than one psychological treatment. This meant that 75% of included studies (n=27) were classed as having CBT as a primary psychological treatment, whether or not it had comprised an 'extra' component such as the involvement of a significant other. Twenty-eight studies (78%) had only one control arm and the remaining eight had two. Of those eight, five were then classified as having a primary WLC, two as a primary TAU control and one as a primary 'symptom monitoring' control.

For the thirty-one studies for which data were available, the duration of the primary psychological treatments selected for analysis varied widely between 6 and 160 hours (mean 28.9; SD 34.1); a variation which was not significantly reduced when in-patient treatments were excluded from the analysis (mean 22.79; SD 22.75).

Outcome Measures

In terms of the measures reported within the included studies, on the assumption that sub-scales of measures would be analysed as a measure in their own right, a total of 278 outcome measures employing continuous scales were deemed to provide data that would facilitate the development of benchmarks.

Excluded Studies

Table 4: Excluded studies and reasons for exclusion

	Primary reason for not meeting inclusion criteria:		
Primary author and year of publication	inappropriate data [a]	no psychological treatment	no appropriate control arm
Babu 2007	○	●	○
Bradley 1987	●	○	○
Buckelew 1998	●	○	○
Buhrman 2004	○	●	○
Dworkin 1994	●	○	○
Dworkin 2002	○	●	○
Fairbank 2005	●	○	○
Freeman 2002	●	○	○
Geraets 2004	○	●	○
Hammond 2001	○	●	○
Jensen 1997	○	○	●
Jensen 2005[b]	●	○	○
Kaapa 2006	○	○	●
Kole-Snijders 1999	●	○	○
Leeuw 2008	○	○	●
Lorig 2008	○	●	○
Marhold 2001	●	○	○
McCarberg 1999	●	○	○

Primary author and year of publication	Primary reason for not meeting inclusion criteria: inappropriate data [a]	no psychological treatment	no appropriate control arm
O'Leary 1988	•	◦	◦
Parker 1988	•	◦	◦
Redondo 2004	◦	◦	•
Smeets 2006	•	◦	◦
Spence 1989	•	◦	◦
Strauss 1986	•	◦	◦
Strong 1998	•	◦	◦
Turner-Stokes 2003	◦	◦	•

Note. The primary authors are for the primary references describing these studies: the Fairbank 2005, Geraets 2004, Hammond 2001, Kole-Snijders 1999, Smeets 2006 and Spence 1989 studies were all described in more than one publication.
[a] 'Inappropriate data' means that raw mean pre and post- treatment and control scores and pre treatment and control standard deviations, or data enabling the calculation of pre-treatment standard deviations or pre-post mean treatment and control change scores and pre-treatment standard deviations were not published or provided by the author, and therefore the data would not facilitate the calculation of effect sizes.
[b] Jensen 2005 was a follow up paper, but the original study data were suitable for analysis and included (Jensen 2001).
• = primary reason for exclusion; ◦= not applicable.

Table 4 clearly demonstrates that the vast majority of studies (n=15; 58%) were excluded as they did not contain extractable data, i.e. the published data did not facilitate the calculation of pre-post treatment effect sizes. Fewer studies (n=6; 23%) were excluded as they did not contain a psychological treatment. The remaining five studies (19%) were excluded for not having a control arm which was explicitly specified as a TAU or that would be viable within PMPs (e.g. an education arm); but rather their self-certified 'control' arms were experimental. In contrast, Table 3 indicates that all included studies with more than one psychological treatment had specified at least one of those treatments as a TAU or had a control that was deemed to be viable within PMPs.

DISCUSSION

The results of this study, in addition to further demonstrating the comparability of RCT with PMP data in terms of treatments and patient groups, suggest that the published literature contains a significant enough volume of data to facilitate the production of benchmarks. The

potential external validity is great given not only this comparability but the fact that more than 3000 patients were entered into the thirty-six included RCTs and that 86% also completed post-treatment measures deemed eligible for inclusion. In terms of the data itself, recommendations that data extracted within systematic reviews should be independently checked for accuracy and completeness (Centre for Reviews and Dissemination, 2009) had been followed within the Cochrane review. As this study involved the extraction of additional data, a proportion that had not previously been extracted was independently reviewed by SM. Whilst it is never possible to completely rule out data entry errors, such attempts to ensure the credibility of the results are considered to be a strength of this study.

In order to try and ascertain the extent to which extracted data would be applicable to PMPs, several methodological matters require consideration. The thirty-six included studies, as originally identified within the 2009 Cochrane review, allowed a large volume of appropriate data to be collected and as such, more recently published literature was not sought. An updated literature search may have been advantageous, however, in allowing more recently published trials to be identified, and determining whether it reflected the PMP reports in study 1 of a move towards third-wave treatments of chronic pain. More recent literature would have been unlikely to exert a great influence on the overall results, but may have overcome, to some extent, the limitation discussed in study 1 of comparing historical RCT data with recent PMP data. As mentioned previously, however, the comparison of treatment components rather than broad treatment types in study 1 may have overcome this to an extent, and pragmatically, the decision to not seek more recently published data was justified. However, the potential impact of not updating the literature search on the external validity of the benchmarks means that future research should aim to identify as much of the literature as possible.

In further consideration of the literature searching methods employed in the 2009 Cochrane review, whilst they were rigorous they did not seek to identify unpublished ('grey') literature. A decision was made to not seek the grey literature within this study given that the overall aim of this research was to determine whether the published literature is applicable to clinical settings, and such a focused research question did not necessitate an exploration of the grey literature. However, it is recognised that ignoring the unpublished literature introduces a publication bias and means that treatment effects are likely to be overestimated as non-significant or negative results are less likely to be published. It is therefore generally agreed that systematic reviews should aim to minimise bias by either identifying unpublished

literature or using statistical methods to correct for the bias (Hopewell, Loudon, Clarke, Oxman & Dickersin, 2009). In future, where relative to the research questions, the grey literature should be sought, but in the meantime, the fact that the benchmarks produced may overestimate treatment effects means that those generated should be interpreted with caution, and viewed as something to aim towards rather a minimum standard within PMPs.

One additional factor with particular reference to the inclusion of data is the methods by which primary psychological treatment and control groups were classified. Preliminary data extraction suggested that the number of studies with multiple treatment and control arms warranted further consideration, and ultimately, the classification was pragmatic. This is because it was not feasible to consider benchmarks for every type of treatment and control group reported, but even without pragmatic restrictions, the total volume of data would not have permitted useful analyses. Classification also represented an attempt to ensure that the outcome data from particular treatments would be as meaningful to PMPs as possible, given the data generated by the sample of participants in study 1 concerning primary psychological treatments. However, it is recognised that the decision to prioritise a WLC or the least intensive control that was deemed to be feasible within PMPs may not produce the most meaningful benchmarks to PMPs. The decision to exclude studies based on the control groups they employed may also have been erroneous. As discussed in study 1, however, without an in-depth knowledge of all treatments offered within PMPs it is only possible to produce benchmarks that may approximate to psychological treatments in clinical settings, and until a more comprehensive understanding of treatments for chronic pain in clinical and research settings is reached, such approximations will have to suffice.

In spite of these methodological considerations, the results so far imply that there are enough similarities between clinical settings and the published literature to ensure that benchmarks can be meaningful, and that data within the published literature can facilitate their development. This study has only touched upon the volume of outcome measures reported within the published literature, but the manner in which they can be aggregated within domains in order to produce benchmarks requires further consideration; as explored within study 3.

STUDY 3: ARE THERE PARTICULAR OUTCOME DOMAINS WITHIN WHICH IT IS USEFUL TO PRODUCE BENCHMARKS FOR USE IN CLINICAL SETTINGS?

AIMS

The results of studies 1 and 2 demonstrated that the published literature contains data which would facilitate the development of benchmarks to be meaningfully applied within routine clinical settings. This study aimed to determine whether there are particular outcome domains of utility to clinicians within routine clinical practice, in order that the available data could be aggregated and analysed in a manner useful to PMPs.

BACKGROUND LITERATURE

Authors have utilised a wide variety of psychometrically validated measures within the benchmarking literature. Some have assessed 'global' constructs, such as the Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM) described in Barkham, Gilbert, Connell, Marshall and Twigg (2005) and employed by Barkham Rees et al. (2005) and Barkham et al. (2008). Others have examined specific constructs such as depression (Wade, et al., 1998) using the Beck Depression Inventory (BDI) which was designed by Beck et al. (1961). Client reported indices of positive outcome (Pugh, et al., 2007) and the selection of a primary measure to benchmark have also been described (Weersing & Weisz, 2002), and measures have been grouped by their specificity to the target problem or their reactivity in terms of who they were completed by, in an exploration of moderators of treatment outcome (Minami, et al., 2007).

Within the chronic pain literature, a reduction in pain is often the primary outcome target of pharmacological treatment (Dworkin, et al., 2008). Psychological treatments, however, target a broader scope of constructs, including changes in emotional and physical functioning, coping and pain behaviours. Such outcomes are usually selected on the basis of clinician and researcher preference, and whilst such a selection procedure is important, it unfortunately increases bias as well as heterogeneity between studies (Turk, Dworkin, Revicki, et al., 2008). Bias and heterogeneity present obstacles to comparing efficacy and effectiveness research, but the introduction of common measures or domains aimed to overcome these (Dworkin, et al., 2005; Dworkin, et al., 2008; Thorn, et al., 2007). Accordingly, the core domains of pain intensity, physical functioning, emotional functioning, patient ratings of global improvement and treatment satisfaction, symptoms and adverse events, and participant

disposition have been recommended within clinical trials (Dworkin, et al., 2005; Dworkin, et al., 2008; Turk, et al., 2006). The British Pain Society (2007) also recommend that PMPs measure changes to healthcare use and working status as well as the extent to which beliefs and modes of information processing have normalised; although they have not specified how the latter outcomes can be measured. Additionally, people with multiple chronic pain conditions have suggested the importance of measuring outcome on domains including relationships, sexual and social activities and home and family life (Turk, Dworkin, Revicki, et al., 2008).

The chronic pain literature has demonstrated that in spite of recommendations, researchers, clinicians and people with chronic pain may differ in terms of the extent to which they value particular outcome domains. Such differing opinions are likely to explain why nearly 300 continuous outcome measure scales with benchmarking potential were identified within study 2. However, variation in terms of the levels of importance given to a large number of outcome domains is both a reminder for clinicians and researchers to justify the use of measures within clinical and research settings (Dworkin, et al., 2005), and for this research to defend the production of benchmarks within particular domains.

METHOD

Design

The chosen design of this study was that of the 'Delphi method'. This method involves a structured process within which knowledge and opinion can be gained from participants in a given field in order to facilitate a group judgement (Adler & Ziglio, 1996). Delphi methods generate an expert consensus using an iterative approach within which honest responses can be considered as the researcher co-ordinates discussion but does not share verbatim responses amongst participants. The use of no more than two to three rounds is recommended as diminishing returns are common when more are used (Linstone & Turoff, 2002).

Ethical Clearance

Determining meaningful outcomes to benchmark involved inviting participants of the survey described in study 1 to participate in an electronic study using the Delphi method, and confirmation that ethical approval was not required can be seen in Appendix 1.

Participants: Round One

Typically, the Delphi method involves the selection of experts based on their level of expertise and knowledge. As academic expert opinions on domains to benchmark are already reflected, to an extent, within the published literature, and this thesis aimed to produce benchmarks of use in clinical settings, however, participants were sought amongst clinicians working within PMPs. In total, participants were thirteen members of staff working in PMPs in England; eleven of whom had participated in study 1. The remaining two volunteered to participate following discussions with colleagues who had taken part in study 1. All agreed to form part of an anonymous expert consensus group, which represented the disciplines of clinical psychology (n=3), physiotherapy (n=6) occupational therapy (n=1), nursing (n=2) and rheumatology (n=1).

There was a 92% response rate to the first survey (n=12), of whom 58% were female (n=7) and half were physiotherapists (n=6, one of whom was a senior physiotherapist specialising in pain management and three of whom were clinical specialists). Other participants were clinical psychologists (n=2, 1 of whom was a consultant) and nurses (n=2, one consultant and one specialist in pain) in addition to one consultant rheumatologist and one occupational therapist. The length of time since participants had achieved primary professional qualifications varied between seven and twenty-nine years, with an average of 16.1 years (SD 8.1). The average length of time for which participants had been employed within their current PMP was 5.2 years (range 1.5-10; SD 3.1).

Measures: Round One

In consideration of the recommendations of the British Pain Society (2007) and within the published literature (Dworkin, et al., 2005; Dworkin, et al., 2008; Eccleston, et al., 2009; Turk, et al., 2006), in addition to the results of study 1 and preliminary results of study 2, the first round asked participants to complete a survey which involved them rating the importance of measuring outcomes within particular domains, comment on a generated definition of each domain and suggest specific measures within domains considered to be at least ‘somewhat important’ within PMPs (see Appendix 6). The survey was developed and deployed using the ‘Bristol Online Survey’ method (University of Bristol, 2009) as in study 1.

Procedure: Round One

Participants had initially been invited to take part in this study during study 1. Several months later they were emailed an electronic link to the survey which made up this round. Following a reasonable time frame which gave participants an opportunity to respond to the survey, their responses were summarised using descriptive statistics and a qualitative summary of written responses, and used to develop the survey employed in round two.

Participants: Round Two

Of the thirteen original participants, there was a 69% response rate to the second round (n=9); 56% of whom (n=5) were female and 40% of whom (n=4) were physiotherapists (3 clinical specialists). Three participants (30%) were clinical psychologists, in addition to one nurse and one consultant rheumatologist. Length of time since primary qualification varied between eight and thirty years, with an average of 15.6 years (SD 8.8). The average length of time for which participants had been employed within their current PMP was 6.1 years (range 2-16 years; SD 4.8).

Measures: Round Two

A survey was designed based on participant responses to the first round, which comprised its quantitative and qualitative results (Appendix 7). The survey invited participants to reflect on the collective comments made regarding domain definitions and confirm whether or not they agreed that measures within particular domains should be recorded within PMPs. It should be noted that study 2 was being conducted concurrently to this study, and due to difficulties in assigning some measures in the published literature to a particular domain, participants were also invited to assign the measures that they had suggested in the first survey to an outcome domain.

Procedure: Round Two

The original thirteen participants were emailed an electronic link to the second round of the survey, and invited to respond within a reasonable time frame. Upon completion, all responses were summarised using descriptive statistics and a qualitative summary of written responses. A summary of responses was then sent to participants (Appendix 8) who were also asked for their permission to be acknowledged within this thesis and to indicate whether they would like a copy of the overall thesis results when available.

RESULTS

The results of each round are detailed in Appendices 7 and 8 and so will not be described in detail here. It is important to recall the original aim of this study, however, which was to generate a list of outcome domains based on those reported within the RCT literature, that clinicians agreed were of importance within PMPs. Table 5 displays the level of agreement amongst participants within each round regarding the importance of measuring domains within PMPs.

Table 5: Percentage agreement amongst PMP participants of the measurement of outcome domains

Domain	Round one (n=12)[a]	Round two (n= 10)[b]
Coping and Cognitive Appraisal	92	100
Disability	66	80
Mood	66	100
Pain behaviour or activity	66	50
Pain experience	25	40
Physiology or fitness	25	60
Social Role Functioning	66	70

[a] Percentage agreement that the domain is *very important* to measure within PMPs
[b] Percentage agreement that the domain should *definitely* be measured within PMPs

Table 5 demonstrates that the agreed level of importance of each domain increased between rounds; with the exception of measures of pain behaviour or activity. Within the first round there had also been a lack of consensus regarding domain definitions and outcome domains to which particular measures could be assigned; reflective of the variation observed within the RCT literature and a recently published review of measures for people with chronic pain (Grimmer-Somers, Vipond, Kumar & Hall, 2009). The aim of the second round was therefore amended slightly in order to develop some clarification over domain definitions and

the assignment of measures to domains, with the intention that the results would inform the manner in which data were aggregated within the meta-analysis.

In round two, the pre-specified criterion for consensus was set at 70%. It is common for researchers to set their own level of consensus (Hsu & Sandford, 2007), and so this was discussed within supervision with SM. Seventy percent was deemed significant enough to reflect consensus amongst the small number of participants, but not so high that it was unattainable. Table 5 demonstrates that a consensus was agreed that measures within the domains of coping and cognitive appraisal, mood, disability and social role functioning should definitely be measured within PMPs. Participants additionally suggested the careful consideration of measures within these domains, recommending that measures of anger and frustration be recorded within the mood domain, and that measures of disability would be of greatest utility if they were related to activities that were personally meaningful to patients. Participants also recommended that 'social role functioning' be renamed 'role functioning'; predicting that patients may distinguish between 'social' i.e. work and non-familial relationships and more general 'roles' i.e. those carried out at home. Whilst the degree of overlap between many domains was observed by a number of participants, most were keen to state that overlap should not preclude measurement.

However, the perceived overlap may have contributed to the lack of consensus that measures within the domains of physiology and fitness, pain behaviour or activity and pain experience should definitely be measured within PMPs. Participants had specifically commented on the difficulties in obtaining objective measures within the physiology and fitness domains, with a general view that whilst physical abilities were important to acknowledge, it would be more meaningful to measure client goals rather than physical abilities per se. Participants also made a distinction between pain behaviours and pain activities and recommended their separate classification, and referred to the differences between self-reported and observed behaviours and the need for a more specific definition. The lack of consensus that pain behaviours or activity should be recorded was therefore assumed to reflect the need for a more robust definition rather than the insignificance of measures that could potentially be included within this domain. Whilst the importance of acknowledging patients' pain experience was implied, many participants stated that its measurement was not meaningful within PMPs given that they did not expect it to change,

with others suggesting a fear of measuring pain in case its lack of change or increase was seen to reflect badly upon the PMP.

There was broad agreement between participants' suggestions as to the allocation of outcome measures compared with my own. There were several discrepancies, however, which were discussed with SM in supervision where final decisions were made.

Implications for Domains to Benchmark

In order to overcome the lack of consensus regarding the inclusion of several outcome domains, the results of this study were considered alongside the published literature. This had recommended that RCTs consider the collection of outcomes of pain intensity, physical functioning, emotional functioning, patient ratings of improvement and satisfaction with treatment, symptoms and adverse events and participant disposition (Dworkin, et al., 2005; Turk & Burwinkle, 2005). The British Pain Society (2007) had additionally recommended healthcare use, working status, changes to beliefs and modes of information processing, and patients had suggested relationships, sexual and social activities and home and family life (Turk, Dworkin, Revicki, et al., 2008). Having considered these recommendations, the results of study 2 and the results of this study, the decision was made to classify measures within the domains of pain experience or behaviour, physical functioning, emotional functioning and coping and cognitive appraisal, and to therefore construct benchmarks within these domains. A brief rationale for these decisions follows.

Pain Experience and Pain Behaviour

The lack of participant consensus regarding the inclusion of measures within the 'pain behaviour or activity' domain was assumed to reflect the broad definition itself, and so the literature was consulted. Dworkin et al. (2005) recommend that a domain of 'pain' should incorporate measures of overt expressions of pain, the use of rescue analgesics and other medical treatments as well as the intensity, quality and experience of pain. Following discussions within supervision, overt expressions of pain, the use of rescue analgesics and other medical treatments were deemed to comprise a distinct category which was named 'pain behaviours', and other measures as suggested by Dworkin et al.(2005) were grouped into a 'pain experience' domain. Though a consensus had not been reached in terms of the measurement of pain experience, this was overruled given suggestions that it should at least be acknowledged within PMPs, and research findings that it is the most consistently reported

outcome domain within RCTs (Eccleston, et al., 2009). In addition, anecdotal reports imply that most patients expect it to be measured within any treatment for chronic pain, and it seemed timely to challenge the assumption that pain experience cannot change as a result of psychological intervention, particularly given that the meta-analysis conducted by Morley et al. (1999) had demonstrated that psychological treatments were associated with a reduction in pain experience.

Physical Functioning

Participants had advocated the inclusion of measures within the domains of disability and role functioning, and Dworkin et al. (2005) suggest that a multifaceted domain named 'physical functioning' can incorporate these domains alongside data regarding activities of daily living, sleep, quality of life and physical abilities; many of which had been recorded within the published literature and deemed of importance to patients within PMPs. A 'physical functioning' domain was therefore agreed upon.

Emotional Functioning

Participants were keen to advocate the inclusion of measures of mood, but congruent with the recommendations of Dworkin et al. (2005), the decision was taken to rename this domain emotional functioning, which would also encapsulate measures of anger, frustration and other emotional states, as suggested by participants within round two.

Coping and Cognitive Appraisal

Researchers within the chronic pain field frequently include additional domains of relevance to the interventions under research (Hoffman, Papas, Chatkoff & Kerns, 2007). In consideration of this, the results of the 2009 Cochrane review of psychological treatments for chronic pain (Eccleston, et al.), consensus amongst participants and the recommendations of Breivik et al. (2008), the decision was made to also record appropriate measures within the domain of coping and cognitive appraisal. These constructs are often viewed as difficult to classify due to the recognition that they may inadvertently measure additional constructs such as adjustment (Jensen, Turner, Romano & Karoly, 1991). Also, their classification as helpful or maladaptive may vary depending on whether a person is seen as 'assimilating' or 'accommodating' in response to chronic pain (Van Damme, Crombez, Goubert & Eccleston, 2009). However, as participants suggested, difficulties in classification should not preclude

measurement, and so it was deemed acceptable to include this domain in order that such measures were captured. Final decisions regarding the classification of outcome measures within domains are displayed in the data extraction code book in Appendix 5.

DISCUSSION

In this study, the views of clinicians working within PMPs were considered alongside the preliminary results of study 2 and the published literature in order to determine whether there are particular outcome domains within which it is useful to produce benchmarks applicable to clinical settings.

The methods employed within this study in order to decide upon these domains give credibility to the results. Minami et al. (2008), for example, recommended that outcomes used to construct benchmarks are determined after the literature has been reviewed and the studies and specific samples have been identified. This study followed these recommendations but also went one step further in seeking clinicians' opinions on the literature and then turning to the literature for guidance to overcome the lack of consensus. Such a circular process demonstrates the 'virtuous circle' described by Barkham and Mellor-Clark (2003), and aimed to bridge one of the gaps between research and clinical practice in attempting to ensure that benchmarks could be developed within clinically meaningful domains.

Additionally, the fact that the surveys were conducted electronically meant that they were widely accessible to a number of potential participants, and the phrasing of the questions in a neutral, non-profession specific or expert manner is likely to have ensured accessibility; regardless of professional background or level of PMP experience. Such accessibility was suggested in that participants represented core professions of varying levels of seniority and with variable levels of experience, and this increases the external validity of the results. Another advantage is that in conducting two rounds, participants were given the opportunity to re-evaluate their responses based on the emerging group consensus, without the issue of diminishing returns becoming pertinent.

However, there are a number of factors which require further consideration, particularly with regard the participants, measures employed and methods of data analysis.

Participants

Whilst several PMP core professions were represented within this study, it could be argued that the small and self-selected sample and unfortunate lack of representation of other

professions limits the extent to which the views can be deemed representative of PMPs. Consequently, the external validity of participant views may be limited, and in future the selection of a sample from a wider range of professionals may be beneficial, although issues of confidentiality and researcher induced bias would require further consideration if individuals were invited to participate rather than volunteering. In the meantime, participants' responses should be considered to provide an indication of the ways in which particular outcome domains are viewed by some professionals within PMPs. However, as responses were not used in isolation to determine domains but considered alongside the results of study 2 and recommendations within the published literature, the credibility of the results is increased.

The decision to not seek input from people with chronic pain also requires some explanation. This was partially influenced by recent research suggesting that outcomes relating to relationships, sexual and social activities and home and family care are of importance (Turk, Dworkin, Revicki, et al., 2008), but also by time constraints; particularly given that this study comprises but one part of this overall thesis. However, the fact that these measures have been rated as important by people with chronic pain justifies their further consideration within both future clinical practice and research.

Outcome Domains

Unfortunately, the data analysis highlighted an oversight within the design of the survey; namely that participants had not been given the opportunity to comment on the inclusion or definition of a 'health and social care domain'. Several measures within this domain had been identified within the preliminary results of study 2, and so possibilities for rectifying the situation were discussed in supervision. It was eventually agreed that the unfortunate lack of participant consultation, together with the fact that measures within this domain had not been reported frequently within the chronic pain literature meant that no further action should be taken to rectify the matter. Instead, the decision to exclude measures within this domain was made, and whilst clearly its inclusion in the first place would have been preferable, this was considered an acceptable solution.

With regard to participants' views about other outcome domains, the decision to not explore patient ratings of improvement and treatment satisfaction, as suggested by Dworkin et al. (2005), also requires some explanation; particularly as 40% of patients with chronic pain are thought to be dissatisfied with treatment effects (Breivik, et al., 2006). The preliminary results

of study 2 demonstrated that such data are rarely reported however, and so it would not have been meaningful to produce benchmarks within these domains. It is noteworthy that Dworkin et al. (2005) developed recommendations intended to cover a wide range of clinical trials which have a predominant focus on pharmacological treatments. The primary aim of these treatments is to produce analgesia; hence their additional suggestion that outcomes relating to symptoms and adverse events are considered. Such measures are rarely reported within trials of psychological treatments of chronic pain however, as demonstrated in study 2, and are unlikely to be recorded within PMPs. It was therefore considered inappropriate to ask participants to consider these domains.

Finally, Dworkin et al. (2005) also recommend the recording of data of participant disposition, including how participants were selected for treatment, when and why they dropped out and reasons for exclusion. Such data is considered important in terms of 'profiling' and because these variables may exert an influence on the outcome of treatment. However, despite that fact that such information has been recorded within previous reviews (Eccleston, et al., 2009), it was not deemed relevant to this thesis given its focus on benchmarking. This is why participant disposition was neither recorded in study 2 nor referred to within this study, although future research may benefit from a consideration of such variables, regardless of its primary focus.

Data Analysis

The final limitations of this study concern the methods of data analysis, which could be viewed as a superficial analysis of participant data. Whilst the descriptive statistics were deemed an adequate method of summarising and describing the group consensus, the qualitative data may have benefitted from a more robust method of analysis than the broad summaries which were provided; particularly as they did not attempt to reduce bias. Linstone and Turoff (2002) recommend the use of a validated theoretical framework to assist with qualitative data analysis when an extensive number of open-ended questions are incorporated within the Delphi method, but as participant responses were summarised and described in a rudimentary manner, the results need to be interpreted with caution. Nevertheless, the intention of this study was to determine whether it would be useful to benchmark outcomes within particular domains, and use the views of several clinicians to inform this decision, rather than gain a consensus opinion from the majority of PMP staff. As such, the rounds may be

considered as comprising a pilot study which has reflected a number of opinions regarding some aspects of outcome measurement. The study has also, however, highlighted the need for further consideration of these issues within clinical settings and more extensive and scientifically stringent research. Such issues would require more time, and within this study, this would also have permitted the follow-up of participants to determine whether their participation in the study and its results impacted on the measures used within their PMPs. The wider impact of the discussion of outcome measure domains in comparison with the published literature's recommendations may also therefore benefit from consideration in the future.

STUDY 4: WHAT DO THE BENCHMARKS SUGGEST ABOUT THE IMPACT OF PSYCHOLOGICAL TREATMENTS OF CHRONIC PAIN WITHIN PARTICULAR OUTCOME DOMAINS?

AIMS

The results of study 1 implied that benchmarks of outcomes for psychological treatments of chronic pain, developed from the published literature, could be meaningfully applied to PMPs in routine clinical settings, and study 2 that the published literature contained data which would facilitate their development. Study 3 identified a number of clinically meaningful outcome domains within which benchmarks could be produced. Together, these results suggest that it would be both feasible and meaningful to produce benchmarks for psychological treatments of chronic pain.

The aim of this study was therefore to develop benchmarks which would reflect the impact of psychological treatments of chronic pain within research settings and could therefore give some indication of the relative outcome of psychological treatments within clinical settings.

BACKGROUND LITERATURE

Quintana and Minami (2006) acknowledged that 'the basis of scientific knowledge is the accumulation of research' and that 'to make sense of a body of research, some form of synthesis is needed' (p. 839). Whether narrative or systematic methods should form such synthesis were discussed within study 2, and the decision was made to use studies identified within the systematic Cochrane review conducted by Eccleston et al. (2009) in order to reduce the subjectivity with which the results could be interpreted. Reducing subjectivity is facilitated by the application of statistical procedures to extracted data, as within meta-analysis, which has been defined as 'the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings' (p. 3, Glass, 1976).

Meta-analysis is undoubtedly a pragmatic way in which large volumes of primary data can be analysed for the purposes of benchmarking, allowing multiple overall effects of treatment to be estimated (Glass, 1976; Lipsey & Wilson, 2001; Quintana & Minami, 2006). The considerable statistical power that meta-analyses have in comparison to individual primary studies is also a strength of its design (Centre for Reviews and Dissemination, 2001; Lipsey & Wilson, 2001; Quintana & Minami, 2006).

To date, whilst a large volume of primary literature describing the outcomes of psychological treatments of chronic pain and control groups has been meta-analysed (Eccleston et al., 2009; Morley et al., 1999), the methods have not yet attempted to bridge the gap between research and clinical settings. In attempting to bridge this gap, properties inherent to the psychological treatment of chronic pain literature require consideration; as briefly discussed in the introduction to this thesis. One matter for consideration relates to the heterogeneity of psychological treatments, for although CBT's predominance within research and clinical settings was identified in studies 1 and 2, there is emerging evidence into the application of other psychological treatments in both settings. This raises the question of whether the aggregation of data from different psychological treatments would be meaningful and produce more representative benchmarks. These methods were recommended by Minami et al. (2008), and have been followed by a number of authors (Barkham, Rees, et al., 2005; Barkham, et al., 2008; Minami, et al., 2007). They also seem plausible within the chronic pain literature, given the meta-analytic findings of equivalent effects of psychological therapies (Wampold, 2001; Wampold, et al., 1997). Similarly, the findings within study 1 that PMPs do not select patients on the basis of diagnostic group or site of pain, congruent with the literature's tendency to categorise by presentation to services than diagnosis (Morley, 2008; Morley & Eccleston, 2008) means that the aggregation of data from studies with different diagnostic samples is defensible as an initial enquiry into the feasibility of benchmarking in this literature.

In ensuring that benchmarks are of utility within clinical settings, the manner in which outcome data is meta-analysed also needs to be considered. The extant benchmarking literature described the relative efficacy of treatments to control groups (Weersing & Weisz, 2002) in order that factors such as spontaneous remission be ruled out. Such methods are also recommended to reduce the likelihood that benchmarks overestimate treatment effects (Minami, et al., 2007). Other methods of benchmarking were varied, and determined by the types of data available from related outcome measures. The results of study 2 demonstrated that data from 278 outcome measures on continuous scales was extracted from the included studies; with the potential to facilitate the development of benchmarks using effect sizes and their CIs, as employed in the Scheeres et al. (2008) study. Such methods are highly recommended by several meta-analytic researchers (Becker, 1988; Hedges & Olkin, 1985) due to the growing consensus that effect sizes, as standardised scores, provide the most

meaningful conclusions about aggregated research findings (Henson, 2006; Quintana & Minami, 2006). Effect sizes and their CIs are also suitable when the treatment effects of studies with heterogeneous outcome measures and sample sizes are employed (Centre for Reviews and Dissemination, 2001, 2009; Hedges & Olkin, 1985). Effect sizes suggest the magnitude of a treatment effect within a population sample by comparing pre and post-treatment scores (Dunlap, Cortina, Vaslow & Burke, 1996) and therefore indicate the impact of that treatment on a specified outcome domain within a wider population (Henson, 2006; Quintana & Minami, 2006). CIs indicate the relative precision of an effect size in providing an estimate of the range within which the true population effect size is expected to lie (Lipsey & Wilson, 2001); and so a 95% CI demonstrates, with 95% confidence, that the true population effect size lies within the given range (Dunlap, et al., 1996; Quintana & Minami, 2006).

METHOD

Design

A meta-analytic design was used to address the research question within this study, to enable data from the included studies to be analysed collectively for the purposes of benchmarking.

Ethical Clearance

This study involved the aggregation of data obtained in study 2 within the domains identified in study 3 and therefore did not require the approval of an ethics committee.

Procedure

The meta-analytic literature does not offer consistent guidance on the methods and statistical procedures most appropriate for studies of particular designs, but acknowledges that when studies employing independent-groups pretest-posttest designs are used, then effect sizes can be computed using the metric for independent groups or repeated measure designs (Morris & De Shon, 2002). As the primary concern of this research was pre-post change within treatment groups, the metrics employed were based on those for repeated measure designs; although those for independent group designs would have also been appropriate. The extraction and classification of data from included studies was outlined in the methods section of study 2, which also described the means by which a single SPSS file containing appropriate

data from primary psychological and control arms had evolved. Following the production of this file, the meta-analysis was conducted. The methods were based on those described by Minami et al. (2008) and Lipsey and Wilson (2001) but were adapted for the chronic pain literature, and comprised the following steps:

1. Each included measure was coded by domain (coping and cognitive appraisal, emotional functioning, pain behaviour, pain experience or physical functioning) as decided within study 3.
2. Table 3 (study 2) was examined and used to inform the decision making process about which data to meta-analyse, given the potential for a large number of effect sizes to be generated. Sixteen of the included studies (44%) had reported one psychological treatment and control arm, twenty-one (58%) had only one psychological treatment arm, and twenty-eight (78%) only one control arm. The number of studies reporting data from more than one treatment or control arm meant that a decision had to be made about how to conduct the analysis within the remit of this thesis, whilst also ensuring that the benchmarks were as meaningful to PMPs as possible. The decision was made to select the psychological treatment and control arms that had been classified as 'primary' within study 2, and as only ten of the primary psychological treatments were non-CBT, to conduct meta-analyses of data within two broad categories: all psychological treatments (N=36), and all CBT treatments (n=26). This led to four pairs of comparisons; all psychological treatments with an active (n=20) or WLC (n=16), and all studies with CBT as the primary psychological treatment and an active (n=16) or WLC (n=10).
3. Several studies required corrections to be made prior to effect size computations. In one study, the raw pre and post-treatment scores had not been provided, so the effect size was calculated indirectly using the reported post-treatment minus pre-treatment change score (Bliokas, Cartmill & Nagy, 2007). Data from two studies had to be pooled before the calculations could be applied, either as data had been grouped by gender (Jensen, Bergström, Ljungquist, Bodin & Nygren, 2001) or by history of depression (Zautra et al., 2008), and this was carried out using the equation:

   (1)

$$\boldsymbol{Pooled\ M} = \frac{(\boldsymbol{M_1} \times \boldsymbol{n_1}) + (\boldsymbol{M_2} \times \boldsymbol{n_2})}{\boldsymbol{n_1} + \boldsymbol{n_2}}$$

where ***Pooled M*** is the pooled pre or post treatment score (e.g. for both genders combined), ***M*** is the original pre or post treatment score per group to be pooled (e.g. $\boldsymbol{M_1}$ **=** male, $\boldsymbol{M_2}$ = female) and ***n*** is the sample size within each group. The pooled SD was then calculated using the equation:

(2)

$$\boldsymbol{Pooled\ SD\ =\ \sqrt{\frac{(SD_1{}^2(n_1-1))\ +\ (SD_2{}^2(n_2\ -\ 1))}{(n_1\ +\ n_2)\ -\ 2}}}$$

As the pre-treatment SD was not provided in the Zautra et al. (2008) study, it was estimated using the equation:

(3)

$$\mathbf{SD\ =\ SE\sqrt{N}}$$

where **SE** is the standard error of the mean and **N** is the pre-treatment sample size. In the absence of the actual pre-treatment SD, this formula was considered to provide a 'good enough' estimate.

4. For each of the 278 outcome measure scales, a biased pre-post effect size (Cohen's ***d***) was then calculated, as is considered an appropriate measure when comparisons are to be made between groups (Centre for Reviews and Dissemination, 2001, 2009; Morris & De Shon, 2002; Quintana & Minami, 2006). This was calculated using the equation:

(4)

$$\boldsymbol{d\ =\ \frac{Mpost\ -\ Mpre}{SDpre}}$$

where **Mpost** is the post-treatment mean of the measure, **Mpre** is the pre-treatment mean of the measure and **SDpre** is the pre-treatment SD of the measure. The pre-treatment SD is recommended when effect sizes are being calculated from studies employing independent groups pretest-posttest designs (Becker, 1988). Whilst some authors recommend that the denominator in an equation used to calculate an effect size is the post-test or pooled pre-post SD (Dunlap, et al., 1996), pre-test SDs are more likely to be homogeneous across studies as they have not been influenced by repeated testing

and experimental manipulations, and so are more likely to produce a conservative estimate of effect size.

5. An unbiased pre-post effect size ($\boldsymbol{d_i}$) was then computed for each outcome measure scale so that a positive $\boldsymbol{d_i}$ indicated pre-post improvement and a negative $\boldsymbol{d_i}$ indicated pre-post deterioration. The equation used was:

(5)

$$\boldsymbol{d_i = (1 - \frac{3}{4n_i - 5}) \frac{Mpost - Mpre}{SDpre}}$$

where $\boldsymbol{n_i}$ is the sample size. This equation corrects for small sample sizes in repeated measures studies (Minami, et al., 2008).

6. In order to determine whether there were any outliers that might impact on aggregated effect size results, box and whisker plots were then produced per outcome measure within each of the five specified outcome domains. This is a recommended graphing technique for determining outliers in an effect size distribution (Hedges & Olkin, 1985; Lipsey & Wilson, 2001), and identified twenty-nine outlying effect sizes. Further examination revealed that twenty-eight of these were unlikely to have a disproportionate impact on the overall results. The remaining outlier had been generated from the CBT arm of a study with a WLC within the coping and cognitive appraisal domain (Woods & Asmundson, 2008). The unbiased $\boldsymbol{d_i}$ for this measure was 4.23, compared with a mean unbiased $\boldsymbol{d_i}$ for all measures within this domain amongst all treatments with WLCs of 0.36 (SD 0.84). As this reported unbiased $\boldsymbol{d_i}$ was more than three SDs from the mean, its inclusion was discussed in supervision with SM where the decision was made to monitor its impact on aggregated effect sizes and to then decide whether or not its inclusion, exclusion or correction would be most appropriate.

7. It was common for studies to report multiple measures within a single outcome domain. This poses problems for the assumption of independence fundamental to most forms of statistical analysis, as entering all generated effect sizes into an analysis would overestimate treatment effects by suggesting that the results came from a larger sample. For example, generating effect sizes from a sample of twenty participants who completed five measures each within the emotional functioning domain would erroneously suggest an overall sample of one-hundred independent participants rather

than twenty participants who had completed five dependent measures. In order to reduce the potential bias that this method would have, a composite effect size was generated using the 'aggregate' methods described by Lipsey and Wilson (2001). This ensured that each study contributed a single effect size per treatment group per outcome domain and allowed for the calculation of a mean post-treatment sample size per outcome domain, as required within later computations. Such methods were congruent with those employed by previous meta-analytic researchers (Dworkin, et al., 2005; Dworkin, et al., 2008; Turk, et al., 2006).

8. As per standard meta-analytic procedures, the variance of each $\boldsymbol{d_i}$ value per domain was then computed using the equation:

(6)

$$\boldsymbol{\sigma^2_{d_i}} = \frac{\boldsymbol{2(1 - r_i)}}{\boldsymbol{n_i}} + \frac{\boldsymbol{d_i^2}}{\boldsymbol{2n_i}}$$

where $\boldsymbol{r_i}$ is an estimated value of the correlation between the pre-treatment and post-treatment scores of the outcome measure (Becker, 1988; Minami, et al., 2008). It has been recognised that pre-post treatment effect sizes may be over-inflated if the inherent larger correlations between the pre-post outcomes (when compared with independent group design correlations) are ignored. Taking a pre-post correlation estimate into account means that the final effect size produced is less likely to be a biased overestimate (Dunlap, et al., 1996). As 278 measures had been employed within the included studies, the calculation of individual pre-post correlation estimates was not possible, and thus a reasonable estimate was sought; as recommended by Minami et al. (2008). The recent evaluation of a CBT informed PMP had allowed pre-post correlations for nine outcome measures to be computed; all of which had been completed by between 720 and 833 participants (Morley, et al., 2008). Each of these measures had been reported by several of the included studies (average pain intensity, distress and interference as measured on numerical scales; the BDI; the HADS anxiety and depression subscales; the CSQ catastrophising subscale; the PSEQ (Bandura, 1997) and a five-minute walk test). The average pre-post correlation on these nine measures was 0.53 and hence an $\boldsymbol{r_i}$ value of 0.53 was considered to be a reasonable estimate for application within equation 6. However, as larger estimates of the correlation are known to produce more

conservative estimates of effect size benchmarks (Minami, et al., 2008), estimated $\boldsymbol{r_i}$ values of 0.3 and 0.7 were also used within the following calculations to determine their impact on the overall effect sizes generated.

9. The 'MEANES.SPS' macro, as described by Lipsey and Wilson (2001) was then applied to the four pairs of comparisons (all psychological treatments with an active or WLC and all studies with CBT as the primary psychological treatment and an active or WLC). This generated a standardised mean effect size per domain per comparison, with associated CIs demonstrating the precision of the estimate. The macro ensured that a number of fundamental components of meta-analysis were conducted. For example, the inverse of the variance per domain per study was 'weighted' so that more precise estimates of effect size (usually from larger samples) would have a greater influence on the aggregated standardised mean effect size; which is therefore less biased and more reliable with less sampling error (Becker, 1988; Hedges & Olkin, 1985; Lipsey & Wilson, 2001). It is good practice to report the magnitude of an effect alongside its statistical significance, i.e. the likelihood that it occurred by chance (Feise, 2002), and the macro also allowed statistical significance to be computed in terms of a *p* value. The level of homogeneity amongst the effect sizes contributing to the overall mean effect size was also determined using the *Q* statistic. Homogeneous effect sizes indicate that contributing effect size estimates are similar, and assume that the only variability between studies is at the participant level and is no more than would be expected by chance. Heterogeneous effect sizes indicate variability over and above that at the participant level and this requires consideration within meta-analysis. The preliminary analysis of the data within study 2 had demonstrated the variability amongst included studies in terms of size, specified treatment components, patient groups, and outcome measures employed, and so heterogeneity in terms of outcome was expected. Such heterogeneity indicates random variability at the study and individual participant level, and in these cases, the 'random effects model' of meta-analysis is required in order that each effect size is weighted according to sample size variability, in addition to variation between participants and studies (Centre for Reviews and Dissemination, 2001, 2009; Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Minami, et al., 2008). In practice, this meant that aggregated effect sizes were generated using the 'fixed-effects' model when contributing effect sizes were considered to be statistically homogeneous, and the

'random-effects' model when heterogeneity was observed amongst contributing effect sizes.

RESULTS

Summary of Available Data

As stated in study 2, a total of thirty-eight papers covering thirty-six separate studies met inclusion criteria and contained data which were suitable for analysis. All studies had been published between 1985 and 2008, and a total of 3681 participants had been entered into the studies with a mean number of entered participants per study of 102.25 (SD 81.13)

Participants

Once decisions had been made regarding the primary psychological treatment and control group, it was possible to calculate the number of participants within each treatment or control arm that had been included in the analyses. This demonstrated that a total of 2605 participants within primary psychological treatment or control arms completed pre-treatment measures, with an average of 72.4 participants per study (range 21-469; SD 77.3) and 36.2 per primary psychological treatment or control arm (range 10-312; SD 40.6). It is of note that the average number of participants post-treatment per study was 68.1 (SD 66); and per primary psychological treatment or control arm was 34.1 (SD 36.9), reflecting a crude pre-post drop-out rate of 6%.

Measures

Study 2 demonstrated that each measure reported by individual studies, and, if applicable, its sub-scales, was classed separately and the thirty-six studies reported a total of 278 outcome measures for which effect sizes were computable. In terms of outcome domains, approximately one-third of all outcomes measured physical functioning (35%; n=97), a slightly smaller proportion measured coping and cognitive appraisal (30%; n=84), approximately 15% measured constructs within the domains of pain and emotional functioning (n= 43 and 41 respectively) and 5% (n=13) measured pain behaviour. Studies contributed between one and sixteen individual measures each, with a mean of 9.1 (SD 4.0) per study. Thirty-four of the included studies (94%) reported that at least half of the outcome measures employed had been patient self-ratings, with only two studies administering all measures by a blinded-

researcher (Haldorsen, Kronholm, Skouen & Ursin, 1998; Morone, Greco & Weiner, 2008). Nine studies (25%) had administered at least some measures by a blinded-researcher, 16% (n=6) by a non-blinded researcher, and three studies (8%) reported at least some measures completed by a spouse or family member.

Table 6 demonstrates that 92% of studies reported outcomes within the physical functioning domain (n=33), 86% the pain experience domain (n=31), 81% the emotional functioning domain (n=29), 73% the coping and cognitive appraisal domain (n=26), and 36% the pain behaviour domain (n=13).

Table 6: Primary treatment, control and outcome domain characteristics of included studies

		Primary treatment and control characteristics			Outcome domain				
Primary author and year of publication	N*	Duration (hours)	Treatment	Control	Coping and cognitive appraisal	Emotional functioning	Physical functioning	Pain experience	Pain behaviour
Altmaier 1992	45	nk	CBT [a]	TAU	●	●	●	●	○
Astin 2003	128	20	Mind	Edu	○	●	●	●	○
Basler 1997	94	30	CBT	TAU	●	○	●	●	●
Becker 2000	166	nk	CBT	WLC	○	●	●	●	○
Bliokas 2007	143	66.5	CBT	WLC	●	●	●	●	○
Carson 2005	43	12	LKM	TAU	○	●	○	●	○
Cook 1998	28	10	CBT	Edu	○	●	●	●	●
Ersek 2003	45	10.5	CBT	Edu	●	●	●	●	○
Ersek 2008	256	10.5	CBT [a]	Edu	●	●	●	○	○
Evers 2002	64	11	CBT	TAU	●	●	●	●	○
Flor 1993	120	8	CBT	TAU	●	●	●	●	●
Greco 2004	92	nk	CBT-ex	TAU	●	●	●	○	●
Haldorsen 1998	469	120	CBT	TAU	○	●	●	●	○
Jensen 2001	214	54	CBT	TAU	○	●	●	●	○

		Primary treatment and control characteristics			Outcome domain				
Primary author and year of publication	N*	Duration (hours)	Treatment	Control	Coping and cognitive appraisal	Emotional functioning	Physical functioning	Pain experience	Pain behaviour
Johansson 1998	42	160	CBT [a]	WLC	●	○	●	●	○
Keefe 1990	99	15	CBT	TAU	○	●	●	○	●
Keefe 1996	88	20	CBT-ex	Edu	●	●	●	○	●
Kraaimaat 1995	77	20	CBT	WLC	●	●	●	●	○
Linton 2008	34	nk	BT	WLC	●	○	●	●	○
Mishra 2000	94	18	CBT	TAU	○	○	○	●	○
Moore 1985	43	32	CBT-ex	WLC	○	●	●	●	●
Morone 2008	37	12	Mind	WLC	●	●	●	●	○
Newton-John 1995	44	16	CBT	WLC	●	●	●	●	○
Nicassio 1997	86	15	BT	Edu	●	●	●	●	●
Puder 1988	69	20	CBT	WLC	●	○	●	●	○
Radojevic 1992	59	6	CBT-ex	TAU	●	●	●	○	●
Thieme 2003	61	75	BT [a]	TAU	●	●	●	●	●
Turner 1988	81	16	CBT	WLC	●	○	●	●	●
Turner 1990	96	16	BT-ex	WLC	○	●	●	●	●
Turner 1993	102	12	CT	WLC	●	●	●	●	●

		Primary treatment and control characteristics			Outcome domain				
Primary author and year of publication	N*	Duration (hours)	Treatment	Control	Coping and cognitive appraisal	Emotional functioning	Physical functioning	Pain experience	Pain behaviour
Turner 2006	158	nk	CBT	Edu	●	●	●	●	○
Vlaeyen 1995	71	16.5	CT	WLC	●	●	●	●	○
Vlaeyen 1996	125	18	CT	WLC	●	○	○	●	○
Williams 1996	121	28	CBT	WLC	●	●	●	●	○
Woods 2008	44	12	CBT	WLC	●	●	●	●	○
Zautra 2008	143	16	CBT	Edu	●	●	●	●	○

Note: *N is pre-intervention and does not necessarily match with the number of participants providing data for analysis; [a] treatments provided within in-patient settings; • = measured; ○ = not measured; BT = Behavioural Therapy; BT-ex = Behavioural Therapy plus extra components (Turner 1990 = additional exercise components); CBT = Cognitive Behavioural Therapy; CBT-ex = CBT plus extra components (Greco 2004 = Biofeedback; Keefe 1996, Moore 1985 and Radojevic 1992= CBT with the involvement of significant others); CT = Cognitive Therapy; Edu: Education; LKM = Loving Kindness Meditation; Mind = Mindfulness; nk = not known; TAU = Treatment As Usual; WLC = Waiting List Control.

Treatments

The proportion of studies reporting data from more than one treatment and control arm was described in study 2 and so will not be repeated here.

Table 6 demonstrates that 72% (n=26) of the primary psychological treatment arms were classed as some form of CBT, 11% (n=4) as behaviour therapy, 8% (n=3) as cognitive therapy, 6% (n=2) as mindfulness and 3% (n=1) as loving kindness meditation. The small number of non-CBT treatments had influenced the decision to not construct benchmarks from these studies alone, given their reduced collective power and therefore consequent predicted utility to PMPs.

Nearly half of all included studies (44%; n = 16) contributed data from WLCs to the analyses, with the remaining twenty contributing data from active controls comprising either specified TAU (n=12; 33%) or education components (n=8, 22%).

Aggregated Effect Sizes

Table 7 provides aggregated pre-post treatment and control group effect sizes for all primary psychological treatment groups with active or WLCs, and CBT alone with active or WLCs, and Figure 4 is a graphical representation of each effect size and its associated confidence interval. A total of forty aggregated standardised mean effect sizes were generated; twenty for primary psychological treatment groups and twenty for control groups. The twenty studies with active controls contributed a total of 144 effect sizes (which had been aggregated by domains) to the analysis; thus contributing an average of 7.2 treatment and control effect sizes per study. The sixteen studies with WLCs contributed an average of 7.5 effect sizes each; 120 in total.

Table 7: Pre-post treatment and control group effect sizes by outcome domain

Domain		All treatments with waiting list control		All treatments with active control		CBT with waiting list control		CBT with active control	
		T	C	T	C	T	C	T	C
C & CA	*k*	12	12	13	13	7	7	11	11
	n	276	270	518	483	168	160	442	427
	d	**0.19***	0.08	***0.37****	0.06	0.17*	0.05	***0.37****	0.08
95% CI		-.002, .39	-.08, .23	.23, .51	-.05, .17	-.13, .46	-.10, .21	.2, .54	-.01, .18
EF	*k*	12	12	18	18	7	7	14	14
	n	292	266	917	681	179	158	792	567
	d	***0.30****	0.05*	***0.33****	0.06	0.23*	-0.12	***0.30****	0.06
95% CI		.095, .513	-.11, .22	.22, .43	-.03, .13	-.01, .47	-.28, .03	.19, .41	-.02, .14
P	*k*	16	16	15	15	10	10	11	11
	n	378	363	743	543	252	234	620	431
	d	***0.46****	0.11	***0.50****	**0.24***	***0.41***	0.03	***0.55****	***0.32****
95% CI		.31, .61	-.004, .22	.33, .66	.05, .43	.27, .55	-.10, .15	.37, .75	.12, .51
PB	*k*	4	4	9	9	2	2	7	7
	n	69	70	242	230	35	33	166	174
	d	***0.37***	0.10	***0.29***	**0.14**	0.32	0.06	***0.29***	0.09
95% CI		.06, .67	-.13, .32	.13, .46	.02, .27	-.01, .65	-.27, .40	.07, .50	-.05, .24
PF	*k*	16	16	17	17	10	10	14	14
	n	378	363	876	649	252	234	772	563
	d	***0.37****	0.08	***0.25***	***0.15***	***0.29***	0.002	***0.25****	***0.16***
95% CI		.22, .52	-.02, .18	.16, .34	.07, .22	.17, .42	-.12, .13	.14, .36	.07, .24

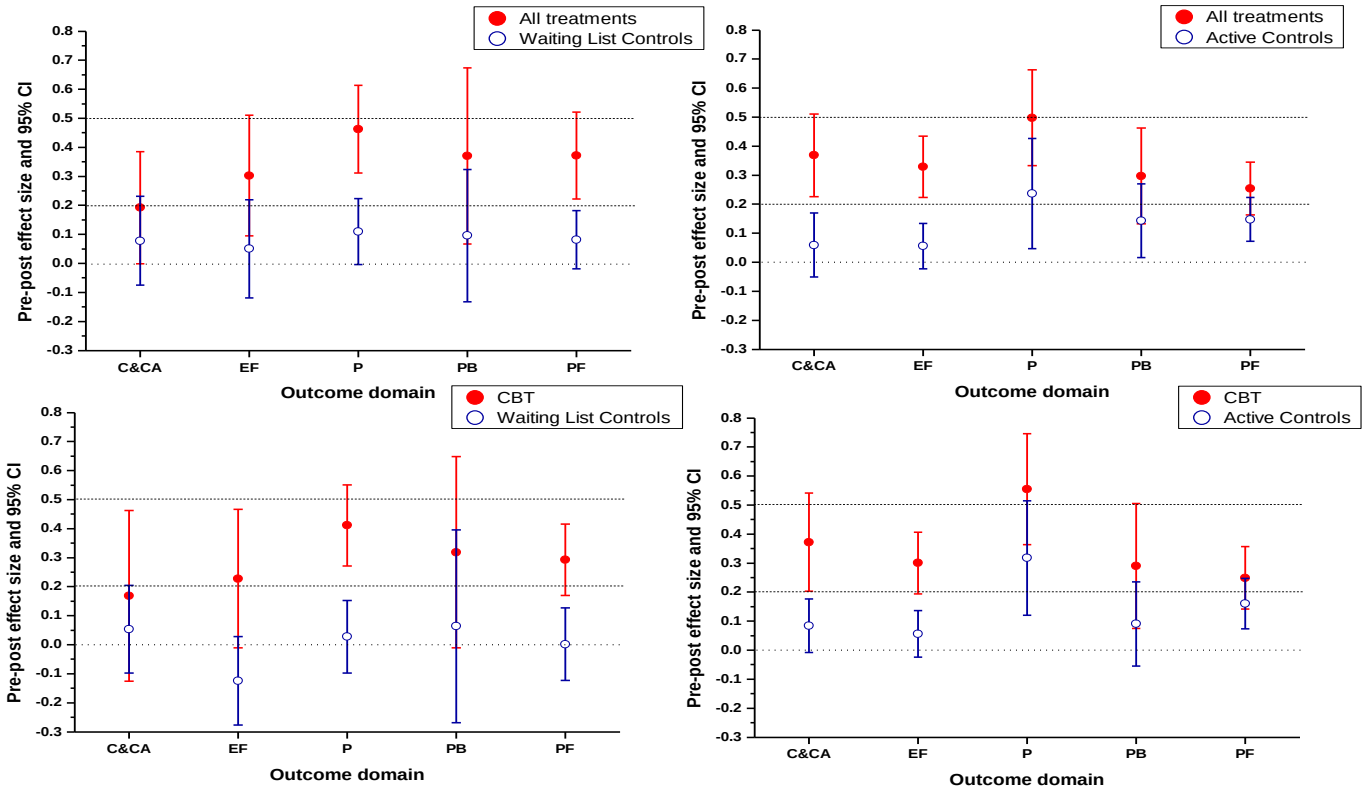
*Note** = computed using the random effects model due to heterogeneous contributing effect sizes.
Effect sizes in bold and italics were significant at the 0.001 significance level. Effect sizes in bold, but not italicised, were significant at the 0.05 significance level. C = control group results; C&CA Coping and cognitive appraisal; CI = confidence interval; EF = Emotional functioning; *d*= standardised mean effect size, computed using the Lipsey & Wilson (2001) macro; *k* = number of included studies entered into the analysis; *n*= average number of participants contributing post-treatment data to each domain; P = Pain experience; PB = Pain behaviour; PF = Physical functioning; T = Treatment group results.

In terms of outcome domains across all treatment and control groups, Table 7 demonstrates that the greatest number of studies and participants contributed data to the meta-analyses within the domain of physical functioning (average 14.25 studies and 510.88 participants respectively). Emotional functioning and pain experience had approximately equivalent contributions in terms of the average number of studies and patients contributing data (12.75 and 13 studies, and 481.5 and 445.5 patients respectively). Fewer studies contributed data within the coping and cognitive appraisal domain (10.75 studies and 343 patients) and studies were least likely to report data within the pain behaviour domain, as indicated by the average 5.5 studies and 127.38 participants contributing data.

As Minami et al. (2008) had suggested, increasing the value of $\boldsymbol{r_i}$ to 0.7 generated a slightly more conservative effect size than when the value of 0.53 was input, However, there was no significant impact upon the mean effect sizes when they were reported to two decimal places and so the estimated $\boldsymbol{r_i}$ value of 0.53 was used to generate the final effect sizes. Analyses had also demonstrated that the outlier of Woods and Asmundson (2008) did in fact impact on the mean effect sizes produced within the coping and cognitive appraisal domain for treatments with WLCs and so it was excluded from the final analyses; as is conventional when outliers are not believed to represent overall study findings (Lipsey & Wilson, 2001).

All effect sizes contributing to those in Table 7 were aggregated under the fixed-effects model unless they were statistically heterogeneous, in which case they were aggregated under the random-effects model. Those generated using the random-effects model have been marked with an asterisk.

Figure 4: Error-bar chart with effect sizes and 95% Confidence Intervals (CIs) of pre-post change within treatment and control groups



Note: C&CA Coping and cognitive appraisal; EF = Emotional functioning; P = Pain experience; PB = Pain behaviour; PF = Physical functioning. Dotted horizontal reference lines represent no change (effect size = 0) and dashed lines represent small (0.2) and medium (0.5) effect sizes.

Effects of all Psychological Treatments Combined

Table 7 demonstrates that when all primary psychological treatment groups with active or WLCs were combined, statistically significant pre-post improvements were observed on measures within all outcome domains. The greatest improvements indicated were within the pain experience domain whether or not the psychological treatment had been conducted alongside a WLC (p<0.001, d =0.46, CI 0.31-0.61,) or active control (p<0.001, d =0.5, CI 0.33-0.66), and following Cohen's descriptors of magnitude that effect sizes of 0.2, 0.5 and 0.8 or above are considered small, medium and large respectively, such effects can be classified as medium (Cohen, 1988, 1992). The significance levels indicate that there was less than a 1% likelihood that these effects were a result of chance, and so the null hypothesis that the primary psychological treatments were not associated with an improvement on these domains is rejected.

Overall, all psychological treatments were associated with a small to medium pre-post improvement. Figure 4 additionally indicates that all psychological treatments combined were superior to WLCs within the pain experience and physical functioning domains as treatment and control CIs did not overlap. Within the domains of coping and cognitive appraisal and emotional functioning, Figure 4 also demonstrates that combined psychological treatments were superior to active controls, and Table 7 that these effects were between small and medium and statistically significant (p<0.001, d =0.37, CI 0.23-0.51 and p<0.001, d =0.33, CI 0.22-0.43 respectively).

Effects of CBT

Figure 4 indicate CBT's superiority to WLCs on the domains of pain experience and physical functioning; which were statistically significant improvements with small to medium effect sizes (p<0.001, d =0.41, CI 0.27-0.55; p<0.001, d =0.29, CI 0.17-0.42 respectively). CBT was also superior to active controls in the domains of coping and cognitive appraisal and emotional functioning (p<0.001, d =0.37, CI 0.2-0.54; p<0.001, d =0.30, CI 0.19-0.41).

Precision of Estimates

The precision of estimates of effect are indicated by the width of CIs, with narrower ranges indicating more precise estimates and wider CIs reflecting a wider range within which

the true estimate of effect size may fall. Additionally, when CIs do not include zero, as demonstrated in most of the treatment CIs in Figure 4, the null hypothesis can be rejected and so it can be concluded that the pre-post change was not zero. There were several exceptions to this, as seen in combined treatment groups with a WLC on the coping and cognitive appraisal domain (p<0.05, d = 0.19, CI -0.002 – 0.39), and all CBT with WLC groups on the coping and cognitive appraisal (p =n.s., d = 0.17, CI -0.13 – 0.46), emotional functioning (p =n.s., d = 0.23, CI -0.11 – 0.47) and pain behaviour (p =n.s., d = 0.32, CI -0.01 – 0.65) domains. As only one of these effect sizes was statistically significant, these results may have been due to chance. It is generally accepted that when a confidence interval includes zero, the estimated effect size should be discarded given the chance that the true effect may reflect no change; and the lack of statistical significance of these effect sizes also implies that this would be advisable. This means that the effect sizes generated on these domains are not meaningful, but suggest that zero pre-post change occurred.

Control Group Effects

Table 7 demonstrates a number of statistically significant effects within active control groups. These were observed within the domains of pain experience and physical functioning amongst all treatments with an active control (p<0.05, d =0.24, CI 0.04-0.43; p<0.05, d =0.15, CI 0.07-0.22 respectively) and the same domains amongst CBT with an active control (p<0.05, d =0.32, CI 0.12-0.51 and p<0.05, d =0.16, CI 0.07-0.24 respectively). It is of note that the CIs of these estimates did not include zero, unlike those of all other control group estimates with the exception of combined treatments with an active control on the pain behaviour domain, where the effect, although statistically significant, was less than small (p<0.05, d =0.14, CI 0.02-0.27). Overall, this indicates that these control groups demonstrated true pre-post change.

Homogeneity of Contributing Effect Sizes

Sixteen of the forty effect sizes (40%) were generated using the random-effects model due to the heterogeneity amongst contributing effect sizes. The pain and emotional functioning domains contained particularly heterogeneous effect sizes (five out of eight in each domain were computed using this model), with half of the aggregated coping and cognitive appraisal and two of the aggregated physical functioning effect sizes comprising heterogeneous estimates.

DISCUSSION

This study aimed to develop a number of benchmarks which would indicate the association between psychological treatments of chronic pain reported in the published literature and pre-post outcomes within specified domains. The intention was to be able to apply these benchmarks to routine clinical settings, thus building a bridge between research and practice. The results demonstrate that within the included clinical trials and following Cohen's descriptors of magnitude (1988, 1992) all psychological treatments combined are associated with a small to medium, and statistically significant, effect on outcome within all specified domains. They also demonstrate that in many cases, the pre-post difference was superior to the outcomes within control groups. The results indicate that measures of pain experience and physical functioning are particularly subject to post-treatment improvement when compared with WLCs, that measures within the coping and cognitive appraisal and emotional functioning domains are subject to post-treatment improvement compared to active controls and that the results of CBT alone are comparable to those when all psychological treatments are combined. The results are broadly congruent with previous literature describing the impact of psychological treatments of chronic pain, which have demonstrated pre-post improvements on measures classed within the pain experience and behaviour (Eccleston, et al., 2009; McCracken, et al., 2007; Morley, et al., 1999), emotional and physical functioning (Eccleston, et al., 2009; McCracken, et al., 2007; McCracken, et al., 2005; Morley, et al., 1999) and coping and cognitive appraisal domains (McCracken, et al., 2007; Morley, et al., 1999).

The similarity between effect sizes and their associated CIs, whether of psychological treatments combined or CBT alone, is a likely reflection of the predominance of CBT within the included studies. Psychological treatments combined or CBT alone demonstrated superiority to WLCs on measures of pain experience (p<0.001, d =0.46, CI 0.31-0.61 and p<0.001, d =0.41, CI 0.27-0.55 respectively) and physical functioning (p<0.001, d =0.37, CI 0.22-0.52 and p<0.001, d =0.29, CI 0.17-0.42 respectively). Similarly, all treatments combined or CBT alone demonstrated superiority to active controls within the coping and cognitive appraisal (p<0.001, d =0.37, CI 0.23-0.51 and p<0.001, d =0.37, CI 0.2-0.54 respectively) and emotional functioning domains (p<0.001, d =0.33, CI 0.22-0.43 and p<0.001, d =0.30, CI 0.19-0.41 respectively).

Whilst effect sizes and their associated CIs were employed in order to produce benchmarks, meaningful benchmarks are those generated when there is no chance that the

pre-post treatment change would be zero or is similar to that demonstrated within control groups. This means that technically, benchmarks have been produced for psychological treatments of chronic pain compared to WLCs within the domains of pain experience and physical functioning, and compared to active controls within the domains of coping and cognitive appraisal and emotional functioning. Intuitively, superiority of psychological treatments to WLCs is more understandable than superiority to active controls, and the suggestion that psychological treatments demonstrate superiority to active controls, but not WLCs, in the domains of coping and cognitive appraisal and emotional functioning, is surprising. Rather than discuss the possible explanations for this here, however, ways in which both the data reported within the included studies and methodological components of this study may have had an impact upon the generated effect sizes have been given further consideration.

The external validity of the results of any study are influenced by power, and the fact that 2605 participants completed pre-treatment measures that were included in the meta-analyses demonstrates the overall power of the results; notwithstanding the estimated crude drop-out rate of 6%. This crude rate did not significantly reduce the number of participants providing data for analysis, but it does reflect the fact that the vast majority of studies only reported data for patients who completed outcome measures at post-treatment and therefore had completed the treatment programme. Data from such samples is known to lead to overestimated results of any study, and so authors such as Minami et al. (2008) suggest that when the results of meta-analyses are intended to produce benchmarks that clinical settings can strive towards, only trials reporting intention to treat (ITT) samples are included. Some authors have followed these recommendations and hence have increased the external validity of their results (Minami, et al., 2007; Scheeres, et al., 2008; Wade, et al., 1998). However, within this study, such methods would have led to the exclusion of more than 80% of data from otherwise included studies and so they were not followed. Rather than diminishing the utility of the resulting effect sizes as benchmarks, however, they ought to simply be interpreted with caution and considered by PMPs as guidelines to strive towards. In the future, both research and practice would benefit from the reporting of ITT samples, however, in order that PMPs had more realistic expectations of outcome, and could perhaps also generate completion rate benchmarks, as within the study of Curtis, Ronan, Heiblum and Crellin (2009).

Pain behaviour was the only domain within which no benchmarks were generated, and only 5% of all measures extracted were classed within this domain. Table 7 indicates that all treatment generated effect sizes within this domain were between small and medium, but Figure 4 demonstrates a lack of precision of estimates. Had measures within this domain been reported more frequently, the results may have been different but clearly, in spite of recommendations that overt expressions of pain, the use of rescue analgesics and other medical treatments are recorded within clinical trials (Dworkin, et al., 2005), this is yet to become routine. Future researchers and clinicians therefore need to give greater consideration to the use of outcomes within this domain.

Another data extraction consideration was the observation that SDs of the raw scores used to compute effect sizes were frequently almost as large as the actual mean raw scores. This indicates heterogeneity in terms of participant scores, and is another factor which means that the benchmarks should be interpreted with caution, although given the heterogeneity reported by PMPs in terms of treatments, patients and outcome measures (study 1), variability in outcome scores may also be common within PMPs; and so this heterogeneity may in fact increase the external validity of the benchmarks. With additional regard to data extraction, it is unfortunate that in spite of the collection of follow-up data, time constraints meant that it was not possible to generate pre to follow-up effect size benchmarks. In future, the early classification of primary psychological and control treatment arms would limit the extraction of data and so allow time for such analyses to take place, but regardless of classification, research should strive to make use of any follow-up data reported.

The validity of the results of any mode of research synthesis are limited by the quality of the included studies as well as the methods of synthesis. Whilst meta-analysis was deemed an appropriate method for the synthesis of the chronic pain literature, it has been argued that it is a crude way in which to group individual clients' data (Kazdin, 2008). Slavin (1986) also argues that combining studies of varying methodological quality is an inherent flaw within meta-analyses which can limit their external validity. With this in mind, the potential exclusion of studies that did not meet pre-defined quality standards was considered. The final decision to not use quality as an exclusion criterion, however, was influenced by a number of factors, including an inability to decide whether it would be more useful to exclude on the basis of inadequate study design quality, treatment quality or overall quality; all of which had been described in an objective measure for rating the quality of trials reporting psychological

treatments of chronic pain (Yates, Morley, Eccleston & Williams, 2005). Eccleston et al. (2009) had reported a weak association between study design and treatment quality and a consequent lack of consistency within studies regarding their quality scores. In terms of overall treatment quality, at least 25% of the included studies (n=9) were not considered to have achieved what the authors of the quality scale deemed to be an 'adequate' score, and as such would have been excluded from the analyses within this study. More significant than the reduced power that would have resulted from such exclusion, however, was the recognition that including studies with varying levels of quality would be more likely to produce benchmarks with greater external validity, as treatments provided within PMPs are unlikely to be of equivalent quality; in whatever way it is judged. Additionally, the included studies had met a number of other pre-defined criteria of direct relevance to the research questions, and so to an extent, the suitability of their inclusion had been considered by other means. Future researchers, however, may wish to consider quality and whether it should determine inclusion, as some authors have recommended (Lipsey & Wilson, 2001).

The procedure conducted within this study comprised a complex collection of elements of methods reported within the published literature; none of which were completely appropriate for the chronic pain literature given its many idiosyncrasies. For example, preliminary analyses of the included studies demonstrated the potential for a single study to contribute myriad effect sizes to the meta-analyses, which would have overestimated treatment effects due to a violation of the assumption of statistical independence (Lipsey & Wilson, 2001). A conservative response to this is to ensure that each study only contributes a single effect size (Quintana & Minami, 2006); wherein a great deal of potentially useful data is discarded. A less conservative response is to follow the methods of aggregation across domains and studies, as recommended within the chronic pain literature and employed within the depression literature (Dworkin, et al., 2005; Dworkin, et al., 2008; Minami, et al., 2008; Minami, et al., 2007; Turk, et al., 2006). Aggregation ensured that most of the data that would be of utility to PMPs was used to generate the effect sizes, and data were also weighted to ensure that studies made a fair contribution to the overall results (Lipsey & Wilson, 2001; Minami, et al., 2008; Quintana & Minami, 2006). However, the decision making process leading to the aggregation of particular data may have influenced the results. For instance, study 2 described in detail how primary psychological and control groups were classified in an attempt to ensure that aggregated data would be of as much utility to PMPs as possible. In spite of this,

Figure 4 demonstrates a lack of superiority of psychological treatments to active controls on the domains of pain experience and behaviour and physical functioning. It also displays a more surprising finding that psychological treatments were only associated with superior pre-post change on measures of coping and cognitive appraisal and emotional functioning when compared with an active control. The similarity in treatment and control results may be a consequence of flaws within the process by which active controls were classified, for even though efforts were made to ensure that the 'active' control was the least intensive treatment offered, collectively, they may have been more similar in content to psychological treatments than intended. This has implications for PMPs in terms of the data that they use as benchmarks, and implies that future research may benefit from a more in-depth consideration of the classification of primary psychological treatment and control arms.

With further reference to the potential impact of data aggregation on the results, the use of the random-effects model meant that, in spite of heterogeneity between and within studies, a meaningful and fair aggregation of the data reported could be conducted. One assumption of this model, however, is that the literature contributing to a meta-analysis is randomly selected; an assumption that is usually violated due to publication bias and the necessary focus of any research question (Quintana & Minami, 2006). Unfortunately, weighting studies further enhances this publication bias as smaller studies are given less credence (Centre for Reviews and Dissemination, 2001, 2009). Additionally, without correcting for publication bias, it is likely that effects will be an overestimate (Hopewell, et al., 2009). In spite of these unfortunate consequences, the random-effects model is recognised to be of benefit in producing conservative estimates of aggregated effect size when contributing effect sizes are heterogeneous and when researchers aim to generalise the results of the meta-analysis to other settings (Centre for Reviews and Dissemination, 2001, 2009; Quintana & Minami, 2006). This model was used to compute 40% of the aggregated effect sizes (n=16) within the pain (n=5), emotional functioning (n=5), coping and cognitive appraisal (n=4) and physical functioning (n=2) domains. Whilst it is a strength that the use of this model led to more conservative effect size estimates within these domains, the variability in contributors implies that moderating variables may have influenced the results. Moderators may include factors such as initial severity of chronic pain, treatment modality, length of treatment, whether the sample was ITT and its size, therapist factors, allegiance to the mode of therapy and treatment integrity. The professional backgrounds and levels of training amongst therapists delivering

psychological treatments of chronic pain are known to be heterogeneous (Eccleston, et al., 2009; Morley, et al., 1999) and also may exert a moderating influence on outcome. Factors such as the specificity to the target problem and reactivity of the outcome measures are also potential moderators (Minami, et al., 2007), but specificity is difficult to ascertain within the chronic pain literature given the lack of explicit treatment aims. Only 3% of outcome measures in the included studies were non-patient rated, and as such reactivity would also have been difficult to determine. However, future research may benefit from exploring the impact of potential moderators on effect sizes using weighted multiple regression analyses for continuous variables (Hedges & Pigott, 2004; Minami, et al., 2007; Quintana & Minami, 2006). This is of particular importance, for whilst previous research suggests that therapist factors are unlikely to impact on outcome (Wampold, et al., 1997), factors such as allegiance are known to overestimate effect sizes (Roth & Fonagy, 1996), and until moderators are considered, the impact of the psychological treatments themselves on the generated benchmarks can only be implied.

Until the influence of moderators is ascertained, PMPs should strive towards the benchmarks that were generated within the domains of pain experience, physical functioning, coping and cognitive appraisal and emotional functioning rather than considering them as an absolute minimum which should be achieved; as recommended within the depression literature (Minami, et al., 2007). Though conclusions as to the impact of the psychological treatments themselves on the generated benchmarks cannot be drawn, the results of this study suggest an association between psychological treatments of chronic pain and outcomes within the domains of coping and cognitive appraisal and emotional functioning when compared with active controls, and pain experience and physical functioning when compared with WLCs. The impact of these benchmarks in the context of all four research questions and their potential influence on future research and clinical practice will now be considered.

OVERALL DISCUSSION

It has already been acknowledged that the most meaningful benchmarks are those developed when treatments and outcomes are homogeneous across trials and the clinical settings within which their application is intended, and patient groups are also similar enough to warrant a useful comparison. It was recognised from the beginning that these conditions would not be met within the chronic pain literature or PMPs in general clinical settings, and therefore this thesis sought to ascertain the feasibility of benchmarking whilst addressing the heterogeneity inherent within the chronic pain literature. In doing so, a number of necessary and considered compromises were made within four separate studies.

Study 1 sought to determine the extent of the similarities between the published literature and PMPs with specific regard to team, patient, treatment and outcome characteristics, and therefore whether it would be meaningful to use the efficacy literature to develop benchmarks for use in clinical settings. Some issues with discrepancies in treatment components, the manner in which primary psychological treatments were classified, bias within the PMP sample and a lack of complete understanding of the specific aims and contents of treatments in both settings were identified. Ultimately, however, responses were received from participants within a significant proportion of PMPs in England (41%), and the results implied that there were enough similarities between PMPs and RCTs to warrant the meaningful comparison of benchmarks from the published literature to PMPs.

The specific aim of study 2 was to determine whether the published literature contained data which would facilitate the development of benchmarks. The decision to use literature identified within a rigorous and methodologically sound Cochrane review (Eccleston, et al., 2009) was expected to increase the credibility of the results. Publication bias and its potential association with overestimates of treatment effect, given that it was not corrected for, was recognised, and though the most recent published literature was not sought, the examination of specific treatment components regardless of publication date was thought to overcome this to an extent. Within this study, whether the manner in which the primary psychological treatment and control groups had been classified would or would not influence the external validity of the results was raised once again. However, the results implied that the published literature could facilitate the development of 278 separate benchmarks from the outcome measure scales reported by included studies. In

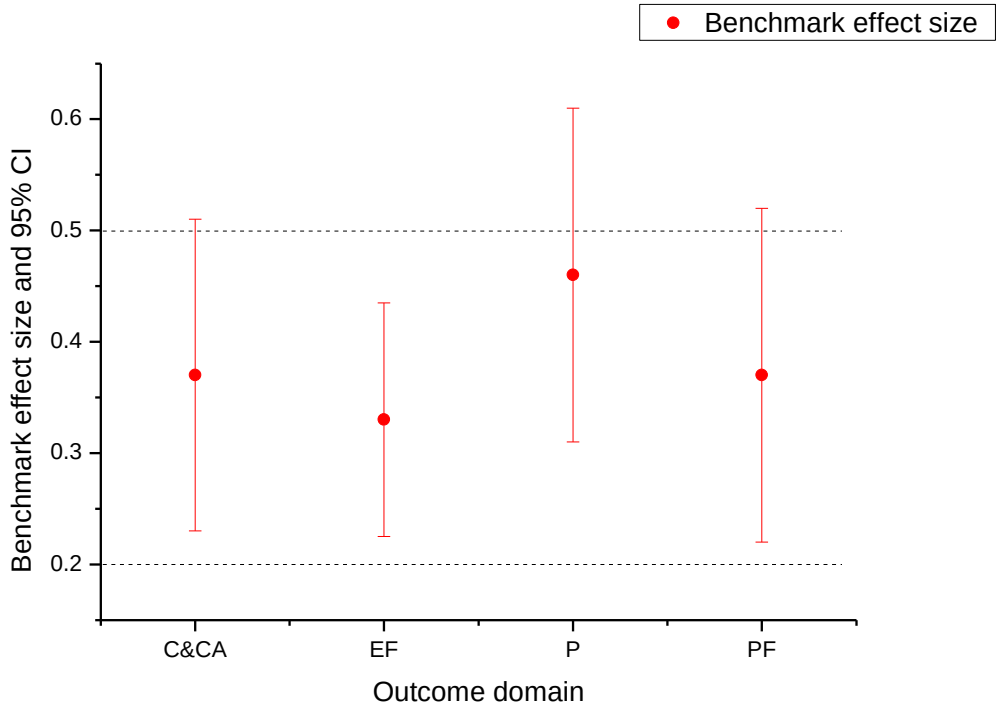
order to use this data in a manner which would be meaningful to PMPs, the categorisation of these scales by outcome domain was then explored within study 3.

In effect, study 3 was a pilot study which aimed to gain a preliminary indication of views of clinicians working in PMPs as to outcome domains to benchmark. The surveys used to obtain their views were widely accessible, although responses did not represent the views of a wide range of professions or of patients themselves. The study raised questions with regard patient ratings of improvement and satisfaction, alongside considerations about the methods used to analyse responses. However, the results were considered alongside highly credible recommendations within the published literature, subsequently leading to the decision to classify measures within the five outcome domains of coping and cognitive appraisal, emotional functioning, physical functioning, pain experience and pain behaviour.

Study 4 then aimed to use the published literature to generate a number of benchmarks within these outcome domains. The methods by which benchmarks were constructed were complex, given that the extant benchmarking literature had not faced issues inherent to the chronic pain literature, therefore could not provide an exact methodological model. However, the methods employed were carefully considered and the power of the results was indicated by the number of patients contributing outcome measure scores to the generated effect sizes. Once again, questions as to the methods of classifying primary psychological treatment and control groups were raised in terms of their potential influence on the benchmarks and ways in which they could be utilised within PMPs. Possibilities that the effect sizes were overestimated due to studies not accounting for attrition or publication bias, or underestimated due to the conservative methods applied to heterogeneous contributing effect sizes were considered, alongside the potential impact of the quality of the trials and the consistency within which outcome measures in specific domains were employed. The decision to aggregate outcome measures within domains was justified given known characteristics of the chronic pain literature, but it was acknowledged that the heterogeneity of the contributing effect sizes may reflect unknown moderator variables; which also may have exerted an impact on the generated effect sizes. In spite of these considerations, the results of study 4 demonstrated the significant contribution that CBT made to the effect sizes generated when psychological treatments were combined, and produced benchmarks for pain experience and physical functioning when compared with WLCs, and coping and cognitive appraisal and emotional functioning when compared with active controls. Figure 5 demonstrates that when

compared to active controls and applying Cohen's descriptors of magnitude (1988, 1992), pre-post improvements within the coping and cognitive appraisal and emotional functioning domains achieved small to medium effects, and when compared to WLCs, pre-post improvements within the pain experience and physical functioning domains are of a similar size. Given the similarity of the generated effect sizes whether psychological treatments were combined or CBT was considered alone, but that combined psychological treatments are likely to be applicable to a greater number of PMPs, the results of combined psychological treatments have been reported.

Figure 5: Benchmarks generated in study 4, with 95% Confidence Intervals (CIs)



Note: C&CA Coping and cognitive appraisal; EF = Emotional functioning; P = Pain experience; PF = Physical functioning. Horizontal lines represent small (0.2) and medium (0.5) effect sizes. C&CA and EF benchmarks were generated from treatments compared with active controls, and P and PF benchmarks from treatments compared with waiting list controls.

Determining the Meaning of the Results

It is of some concern that in spite of these benchmarks, the potential for unexamined variables to have exerted an influence means that it is impossible to conclude whether these results are representative of true effect size estimates, have been overestimated or underestimated. Additionally, in order to fulfil the definition of a benchmark, these evidence-based practice results must be considered in relation to practice-based evidence. In comparing these benchmarks with evidence generated as a result of an adapted practice-based ACT treatment (Vowles & McCracken, 2008) and a CBT

informed PMP (Morley, et al., 2008), several discrepancies were observed. The adapted ACT treatment reported medium to large effects on measures of pain experience, and large effects within the coping and cognitive appraisal, emotional functioning, physical functioning and pain behaviour domains, and the CBT informed PMP demonstrated medium pre-post effects within the domains of emotional functioning, physical functioning and pain behaviour, and medium to large effects of measures within the coping and cognitive appraisal domains. Although this comparison is deliberately crude and for illustrative purposes, the incongruence may indicate that these particular practice-based psychological treatments are more powerful, or of differing levels of intensity and quality to the aggregated data and therefore that the comparison is not a meaningful one. Alternatively, the differences may indicate that the generated benchmarks do not provide a true estimate due to issues previously discussed, or that a combination of both these factors makes the comparison unfair. The reasons for the discrepancies are of less concern, however, than the question of how the benchmarks produced from aggregated RCT data may be of utility within PMPs whilst questions as to their validity remain unanswered. One answer may be that they cannot, for in spite of the suggested similarities between the published literature and routine clinical settings, essentially the chronic pain literature does not lend itself to the production of meaningful benchmarks from meta-analysis; whether or not potentially confounding factors such as the classification of treatments and controls, study quality and moderators are considered. This suggestion is supported by the fact that although, on a practical basis, it has been possible to generate effect sizes and some benchmarks, it is difficult to interpret the results and their magnitude compared with practice-based evidence. This potential answer is of some concern, however, given the intention of this research to bridge a gap between research and practice, and allow PMPs to determine realistic expectations of outcome and the relative effectiveness of their psychological treatments in a clinical governance context (Department of Health, 1999).

The Application of Benchmarks within PMPs

A number of other factors require consideration in relation to the difficulties in determining the meaning of the results, and so how, if at all, they might be applied within PMPs. The aggregation of myriad outcome measures within domains is one such factor, as although it was justified given the heterogeneity of outcomes reported within trials, aggregation makes it difficult to determine the specific meaning of a pre-post change of a given magnitude; whether or not it was in line with practice-based evidence. Rather than aggregating outcome measures, other benchmarking studies have overcome the limitations

of heterogeneous measures by focusing upon particular measures and transforming the scores so that those used in clinical and research settings are comparable (Barkham, Rees, et al., 2005; Barkham, et al., 2008). The focus on particular measures, but not the transformation of scores, would be feasible within the chronic pain literature, and indeed recommendations that clinical pain trials focus on a single measure (Turk, Dworkin, McDermott, et al., 2008) would facilitate this. Whilst such a focus in RCTs would ensure a more straightforward transfer of benchmarks to clinical settings, however, it is unlikely to be seen in the near future, and heterogeneity amongst outcome measures is likely to continue to prevail.

Rather than using a meta-analysis of the relevant literature, other benchmarking studies have avoided the limitations of aggregation by generating benchmarks from several trials providing similar treatments to those in specified clinical settings (Curtis, et al., 2009; Scheeres, et al., 2008; Wade, et al., 1998). This is one option for PMPs, but as it would lead to the development of benchmarks which were meaningful to specific services, it would require a great deal of effort by many PMPs and reduce both the power of the benchmarks and their wider external validity. This thesis deliberately focused on the meta-analysis of data reported within RCTs as such methods are considered to provide the highest quality evidence (Greenhalgh, 2006), not least due to the power of generated results, as the intention was to generate benchmarks of utility to many PMPs. However, that PMPs would benefit more from specific benchmarks generated from several pertinent trials cannot be ruled out. In the meantime, though the 'virtuous circle' described by Barkham et al. (2006) cannot be drawn as a number of unanswered questions prevent the straightforward transfer of benchmarks to PMPs, there are several ways in which they can be applied; alongside the caveats referred to previously.

The transparency of the decision making process around the classification of primary psychological treatment and control arms and the detail provided about each included study means that PMPs might look to the included RCTs and ascertain their comparability to a primary psychological treatment, and therefore how applicable the benchmarks might be. One question that this raises is whether treatments compared with an active control or a WLC would be more applicable to PMPs; a question that is difficult to answer without a comprehensive understanding of all treatments offered within PMPs. However, it is reasonable to conclude that it would be appropriate for PMPs to strive towards the benchmarks produced from combined psychological treatments with a WLC. This option would be viable given that PMPs are more likely to have a WLC group than

another active treatment of any level of intensity, and meaningful given previous findings of equivalent effects of psychological treatments conducted by qualified therapists (Wampold, et al., 1997). Such a decision would mean that only the benchmarks within the domains of pain experience and physical functioning were of use within PMPs, but that WLC benchmarks were also available for comparison, as Table 7 and Figure 4 (study 4) demonstrated that all WLCs had CIs overlapping zero and so imply that no change would be expected within these groups. Study 3 demonstrated that PMP participants reached a consensus that measures within the domains of disability (which were included within the physical functioning domain) should definitely be recorded within clinical settings, but only 40% agreed that outcomes within the domains of pain experience should be measured. It is of note that many participants based their decisions on the assumption that pain experience would not change, yet study 4 demonstrated that change of the greatest magnitude was observed within this domain. The lack of consensus cannot be assumed to reflect actual practices within PMPs though, particularly as study 1 suggested the use of measures within most domains as common practice and the sample of PMPs was small, and so in the meantime, it is reasonable to assume that benchmarks within these domains would be of utility within PMPs.

That the benchmarks, particularly those within the pain experience and physical functioning domains, should be striven towards rather than considered as minimum standards to be achieved has been discussed, and is of particular importance given the inability to conclude whether they have been under or overestimated or the potential impact of moderating variables on outcome. The incongruence between the crude practice-based evidence and evidence-based practice comparisons also suggests that striving would be more appropriate. However, the methods that PMPs may employ to generate effect sizes that can be compared with the benchmarks ought to be addressed.

One suggestion is that clinical settings collect pre-post outcome data within the outcome domain, record the sample size and generate collective means and SDs. Where multiple outcome measures within a domain were recorded, the decision could be made as to whether data from a single measure should be used, or whether an aggregated mean score would be more useful. PMPs would then need to apply the equation described in step 5 of the procedure in study 4, namely:

$$\boldsymbol{d_i} = \left(\mathbf{1} - \frac{\mathbf{3}}{\mathbf{4}\boldsymbol{n_i} - \mathbf{5}}\right)\frac{\mathbf{Mpost} - \mathbf{Mpre}}{\mathbf{SDpre}}$$

Congruent with the suggestion that PMPs strive towards these benchmarks, a simple comparison of effects and their magnitude may suffice. In the event that the methodological constraints preventing conclusions from being drawn are rectified, however, PMPs may choose to employ more precise methods of comparison such as those described by Curtis et al. (2009) and Minami et al. (2009). Both publications detail how this data could be used to calculate a critical value that the clinical effect size must exceed to be classified as clinically equivalent to the benchmark, and essentially this means that a difference of less than 0.2 or 10% would suggest equivalence.

PMPs could also generate CIs and determine the statistical significance of pre-post treatment improvements to compare with benchmarks. This would be possible using the equation reported in step 8 of the procedure of study 4 to estimate the variance of $\boldsymbol{d_i}$, using the $\boldsymbol{r_i}$ value of 0.53 or their own estimates of pre-post correlation of the outcome measure, namely:

$$\boldsymbol{\sigma^2_{d_i}} = \frac{\boldsymbol{2(1 - r_i)}}{\boldsymbol{n_i}} + \frac{\boldsymbol{d_i^2}}{\boldsymbol{2n_i}}$$

Clinical settings have previously used such methods in benchmarking psychological treatments of CFS (Scheeres, et al., 2008) and the precise calculations have been described by Lipsey and Wilson (2001).

Future Considerations

Regardless of the how the generated benchmarks are interpreted and applied within PMPs, this thesis has highlighted a number of factors that require consideration within research and clinical settings to ensure that, in the future, meaningful benchmarks are generated from the chronic pain literature; not least to meet clinical governance requirements (Department of Health, 1999). Following consideration of these factors, such research might achieve some of the aims that were not fully realised within this thesis; including bridging the gap between research and clinical settings, understanding what the benchmarks suggest about the impact of psychological treatments of chronic pain and determining realistic outcomes.

Trials reporting psychological treatments of chronic pain must consider the clinical utility of reported outcome measures. These would influence both the available methods of benchmarking and how clinically meaningful the results were. Similarly, rather than conducting a meta-analysis of many trials, the use of data from several trials of relevance to PMPs may produce more clinically meaningful benchmarks. Homogeneity of standardised outcome measures in particular would mean that the proportion of patients

achieving statistically reliable change (Jacobson & Truax, 1991; Vowles & McCracken, 2008), normative ranges on outcome measures (Barkham, et al., 2008; Scheeres, et al., 2008; Wade, et al., 1998), clinically important change (Dworkin, et al., 2008) or sample size dependent critical cut-off values (Minami, et al., 2008) could be determined. However, whether or not future benchmarking methods involve meta-analyses or comparison with several pertinent trials, the potential impact of moderators needs to be explored in order that the extent of the impact of psychological treatments of chronic pain can be determined. Follow-up data should also be analysed whenever possible to ensure that the stability of any treatment effects over time can be ascertained. Until such time, the impact of psychological treatments of chronic pain on outcomes within specified domains can only be assumed.

To conclude, the issues precluding the straightforward comparison of clinical and research data, including determining whether PMP and efficacy data is comparable, the likely over or underestimation of these benchmarks and the potential impact of moderator variables are unlikely to be resolved in the near future. The overall aim of this thesis was to assess the feasibility of generating benchmarks which would reflect the impact of psychological treatments of chronic pain and be meaningful to routine clinical settings, however, and absence of evidence does not imply evidence of absence. Until the issues discussed are resolved, recommendations that the generated benchmarks be seen as a general standard to strive towards, alongside a consideration of their limitations, is viable. This would ensure their transfer to clinical settings as intended; although that benchmarks developed from several RCTs would be more meaningful to PMPs than those generated from a meta-analysis cannot be ruled out.

REFERENCES

Adams, N., Poole, H. & Richardson, C. (2006). Psychological approaches to chronic pain management: part 1. *Journal of Clinical Nursing, 15*(3), 290-300.

Adler, M. & Ziglio, E. (Eds.). (1996). *Gazing into the Oracle: The Delphi Method and its Application to Social Policy and Public Health*. London: Jessica Kingsley.

APA (2004). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)* (4th ed.). Washington DC: American Psychiatric Association.

Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Worth Publishers.

Barker, C., Pistrang, N. & Elliott, R. (2002). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners* (2nd Edition ed.). Chichester: John Wiley & Sons Ltd.

Barkham, M., Gilbert, N., Connell, J., Marshall, C. & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry, 186*(3), 239-246.

Barkham, M. & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice based evidence: developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology and Psychotherapy, 10*(6), 319-327.

Barkham, M., Mellor-Clark, J., Connell, J. & Cahill, J. (2006). A core approach to practice-based evidence: a brief history of the origins and applications of the CORE-OM and CORE system. *Counselling and Psychotherapy Research, 6*(1), 3-15.

Barkham, M., Rees, A., Leach, C., Shapiro, D., Hardy, G. & Lucock, M. (2005). Rewiring efficacy studies of depression: an empirical test in transforming BDI-I to CORE-OM scores. *Mental Health and Learning Disabilities Research and Practice, 2*(1), 11-18.

Barkham, M., Stiles, W., Connell, J., Twigg, E., Leach, C., Lucock, M., et al. (2008). Effects of psychological therapies in randomized trials and practice-based studies. *British Journal of Clinical Psychology 47*(4), 397-415.

Beck, A., Ward, C., Mendelson, M., Mock, J. & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561-571.

Becker, B. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*(2), 257-278.

Boersma, K. & Linton, S. (2006). Psychological processes underlying the development of a chronic pain problem: a prospective study of the relationship between profiles of psychological variables in the fear-avoidance model and disability. *Clinical Journal of Pain, 22*(2), 160-166.

Breivik, H., Borchgrevink, P., Allen, S., Rosseland, L., Romundstad, L., Kvarstein, G., et al. (2008). Assessment of pain. *British Journal of Anaesthesia, 101*(1), 17-24.

Breivik, H., Collett, B., Ventafridda, V., Cohen, R. & Gallacher, D. (2006). Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *European Journal of Pain, 10*(4), 287-333.

British Pain Society (2007). *Recommended guidelines for Pain Management Programmes for adults*. London: The British Pain Society.

British Pain Society (2009). Pain Management Programme Society SIG Retrieved 24th August 2009, from http://www.britishpainsociety.org/members_sigs_pmp.htm

British Psychological Society (2007). *Committee on Training in Clinical Psychology: Criteria for the accreditation of postgraduate training programmes in clinical psychology*. Leicester: British Psychological Society.

Cano, A., Gillis, M., Heinz, W., Geisser, M. & Foran, H. (2004). Marital functioning, chronic pain, and psychological distress. *Pain, 107*(1-2), 99-106.

Centre for Reviews and Dissemination (2001). *Undertaking systematic reviews of research on effectiveness:*
CRD's guidance for those carrying out or commissioning reviews (2nd Edition). York: University of York.

Centre for Reviews and Dissemination (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York: University of York.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cook, D., Mulrow, C. & Haynes, R. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine 126*(5), 376-380.

Cook, T. & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Company.

Cooper, H. & Hedges, L. (1994). Research synthesis as a scientific enterprise. In H. Cooper & L. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 3-14). New York: Russell Sage Foundation.

Curtis, N., Ronan, K., Heiblum, N. & Crellin, K. (2009). Dissemination and effectiveness of multisystemic treatment in New Zealand: a benchmarking study. *Journal of Family Psychology, 23*(2), 119-129.

Davies, H. & Crombie, I. (2001). What is a Systematic Review? Retrieved 26th January 2009, from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Syst-review.pdf

DeCoster, J. (2004). Meta-analysis Notes Retrieved 12th January 2009, from http://www.stat-help.com/notes

Department of Health (1999). *The National Service Framework for Mental Health*. London: HMSO.

Department of Health (2009). *Essence of Care: a consultation on a new benchmark on pain*. Retrieved 24th May 2010. from http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_103061.pdf.

Donovan, M. I., Evers, K., Jacobs, P. & Mandleblatt, S. (1999). When there is no benchmark: designing a primary care-based chronic pain management programme from the scientific basis up. *Journal of Pain and Symptom Management 18*(1), 38-48.

Dunlap, W., Cortina, J., Vaslow, J. & Burke, M. (1996). Meta-analysis of experiments with matched groups of repeated measures designs. *Psychological Methods, 1*(2), 170-177.

Dworkin, R., Turk, D., Farrar, J., Haythornthwaite, J., Jensen, M., Katz, N., et al. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations *Pain, 113*(1-2), 9-19.

Dworkin, R., Turk, D., Wyrwich, K., Beaton, D., Cleeland, C., Farrar, J., et al. (2008). Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *Journal of Pain, 9*(2), 105-121.

Eccleston, C., Williams, A. & Morley, S. (2009). Psychological therapies for the management of chronic pain (excluding headache) in adults. [Cochrane Review]. Retrieved 11th May 2009, from The Cochrane Library, Wiley Interscience: http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD007407/frame.html

Feise, R. (2002). Do multiple outcome measures require p-value adjustment? *BMC Research Methodology* Retrieved 12th January 2010, from http://www.biomedcentral.com/content/pdf/1471-2288-2-8.pdf

George, S. & Hirsh, A. (2005). Distinguishing patient satisfaction with treatment delivery from treatment effect: a preliminary investigation of patient satisfaction with symptoms after physical therapy treatment of low back pain. *Archives of Physical Medicine & Rehabilitation, 86*(7), 1338-1344.

Glass, G. V. (1976). Primary, secondary and meta-analysis. *Educational Researcher, 5*(10), 3-8.

Goodheart, C., Levant, R., Barlow, D., Carter, J., Davidson, K., Hagglund, K., et al. (2006). Evidence-based practice in psychology: APA presidential task force on evidence-based practice. *American Psychologist, 61*(4), 271-285.

Green, L. & Glasgow, R. (2006). Evaluating the relevance, generalization and applicability of research: Issues in external validation and translation methodology. *Evaluation and the Health Professions, 29*(1), 126-153.

Greenhalgh, T. (2006). *How to read a paper: the basics of evidence based medicine* (3rd Edition ed.). Oxford: Blackwell Publishing.

Grimmer-Somers, K., Vipond, N., Kumar, S. & Hall, G. (2009). A review and critique of assessment instruments for patients with persistent pain. *Journal of Pain Research, 2*, 21-47.

Hedges, L. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.

Hedges, L. & Pigott, T. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*(4), 426-445.

Henson, R. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist, 34*(5), 601-629.

Hirsh, A., Atchison, J., Berger, J., Waxenberg, L., Lafayette-Lucey, A., Bulcourf, B., et al. (2005). Patient satisfaction with treatment for chronic pain: predictors and relationship to compliance. *Clinical Journal of Pain, 21*(4), 302-310.

Hoffman, B., Papas, R., Chatkoff, D. & Kerns, R. (2007). Meta-analysis of psychological interventions for chronic low back pain. *Health Psychology, 26*(1), 1-9.

Hopewell, S., Loudon, K., Clarke, M., Oxman, A. & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results [Methodology

Review]. Retrieved 20th March 2010, from The Cochrane Library, Wiley Interscience: http://mrw.interscience.wiley.com/cochrane/clsysrev/articles/MR000006/frame.html

Hsu, C. & Sandford, B. (2007). The delphi technique: making sense of consensus. *Practical Assessment, Research and Evaluation, 12*(10), 1-8.

Jacobson, S. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19.

Jensen, M., Turner, J., Romano, J. & Karoly, P. (1991). Coping with chronic pain: a critical review of the literature. *Pain, 47*(3), 249-283.

Kabat-Zinn, J., Lipworth, L. & Burney, R. (1985). The clinical use of mindfulness meditation for the self-regulation of chronic pain. *Journal of Behavioral Medicine, 8*(2), 163-190.

Karoly, P. & Ruehlman, L. (2006). Psychological 'resilience' and its correlates in chronic pain: findings from a national community sample. *Pain, 123*(1-2), 90-97.

Kazdin, A. (2008). Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*(3), 146-159.

Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist, 40*(11), 1189-1202.

Kirsch, I. (Ed.). (1999). *How Expectancies Shape Experiences*. Washington DC: American Psychological Association.

Kwan, K., Roberts, L. & Swalm, D. (2005). Sexual dysfunction and chronic pain: the role of psychological variables and impact on quality of life. *European Journal of Pain 9*(6), 643-652.

Linstone, H. & Turoff, M. (2002). The Delphi Method: Techniques and Applications Retrieved 14th January 2010, from http://www.is.njit.edu/pubs/delphibook/index.html#toc

Lipsey, M. & Wilson, D. (2001). *Practical Meta-Analysis* (Vol. 49). London: Sage.

Lynch, M., Campbell, F., Clark, A., Dunbar, M., Goldstein, D., Peng, P., et al. (2008). A systematic review of the effect of waiting for treatment for chronic pain. *Pain, 136*(1-2), 97-116.

McCracken, L., Evon, D. & Karapas, E. (2002). Satisfaction with treatment for chronic pain in a specialty service: preliminary prospective results. *European Journal of Pain, 6*(5), 387-393.

McCracken, L., Gauntlett-Gilbert, J. & Vowles, K. (2007). The role of mindfulness in a contextual cognitive-behavioural analysis of chronic pain related suffering and disability. *Pain, 131*(1-2), 63-69.

McCracken, L., Vowles, K. & Eccleston, C. (2005). Acceptance-based treatment for persons with complex, longstanding chronic pain: a preliminary analysis of treatment outcome in comparison to a waiting phase. *Behaviour Research and Therapy, 43*(10), 1335-1346.

Merskey, H. & Bogduk, N. (1994). *Classification of chronic pain: Syndromes and Definitions of Pain Terms* (2nd ed.). Seattle: International Association for the Study of Pain (IASP Press).

Minami, T., Davies, D., Tierney, S., Bettmann, J., McAward, S., Averill, L., et al. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology, 56*(2), 309-320.

Minami, T., Serlin, R., Wampold, B., Kircher, J. & Brown, G. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality and Quantity, 42*(4), 513-525.

Minami, T., Wampold, B., Serlin, R., Kircher, J. & Brown, G. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology, 75*(2), 232-243.

Morley, S. (2008). Trial design of psychological treatments in chronic pain: what can we tell patients? In H. McQuay, E. Kalso & A. Moore (Eds.), *Systematic Reviews in Pain Research: Methodology Refined* (pp. 217-232). Seattle: International Association for the Study of Pain.

Morley, S. & Eccleston, C. (2008). Cognitive-behavior therapy for chronic pain in adults In J. Haythornthwaite (Ed.), *Textbook of Clinical Pain Management. Chronic Volume* (2nd ed., pp. 131-138). Oxford: Oxford University Press.

Morley, S., Eccleston, C. & Williams, A. (1999). Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. *Pain, 80*(1-2), 1-13.

Morley, S., Williams, A. & Hussain, S. (2008). Estimating the clinical effectiveness of cognitive behavioural therapy in the clinic: evaluation of a CBT informed pain management programme. *Pain, 137*(3), 670-680.

Morris, S. & De Shon, R. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105-125.

NHS Executive (1996). *NHS Psychotherapy Services in England: Review of Strategic Policy*. London: Department of Health.

Philips, H. & Rachman, S. (1996). *The Psychological Management of Chronic Pain: A Treatment Manual*. New York: Springer

Price, D., Milling, L., Kirsch, I., Duff, A., Montgomery, G. & Nicholls, S. (1999). An analysis of factors that contribute to the magnitude of placebo analgesia in an experimental paradigm. *Pain, 83*(2), 147-156.

Pugh, M., Berlowitz, D., Montouris, G., Bokhour, B., Cramer, J., Bohm, V., et al. (2007). What constitutes high quality of care for adults with epilepsy? *Neurology, 69*(21), 2020-2027.

Quintana, S. & Minami, T. (2006). Guidelines for meta-analyses of counseling psychology research. *Counseling Psychologist, 34*(6), 839-877.

Rosensteil, A. & Keefe, F. (1983). The use of coping strategies in chronic low back pain. *Pain, 17*(1), 33-44.

Roth, A. & Fonagy, P. (1996). *What works for whom? A critical review of psychotherapy research*. London: Guilford Press.

Scheeres, K., Wensing, M., Knoop, H. & Bleijenberg, G. (2008). Implementing cognitive behavioral therapy for chronic fatigue syndrome in a mental health center: a benchmarking evaluation. *Journal of Consulting and Clinical Psychology, 76*(1), 163-171.

Seligman, M. (1995). The effectiveness of psychotherapy: the Consumer Reports study. *American Psychologist, 50*(12), 965-974.

Shadish, W., Matt, G., Navarro, A. & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Journal of Consulting and Clinical Psychology, 126*(4), 512-529.

Slavin, R. (1986). Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(1), 5-11.

The Cochrane Collaboration (2010). Cochrane Reviews: What are Cochrane Reviews? Retrieved 20th March, 2010, from http://www.cochrane.org/cochrane-reviews

Thorn, B., Cross, T. & Walker, B. (2007). Meta-analyses and systematic reviews of psychological treatments for chronic pain: relevance to an evidence-based practice. *Health Psychology, 26*(1), 10-12.

Thorne, F. & Morley, S. (2009). Prospective judgements of acceptable outcomes for pain, interference and activity: patient-determined outcome criteria. *Pain, 144*(3), 262-269.

Tunks, E., Crook, J. & Weir, R. (2008). Epidemiology of chronic pain with psychological comorbidity: prevalence, risk, course, and prognosis. *Canadian Journal of Psychiatry, 53*(4), 224-234.

Tunks, E., Weir, R. & Crook, J. (2008). Epidemiologic perspective on chronic pain treatment. *Canadian Journal of Psychiatry, 53*(4), 235-242.

Turk, D. (1999). The role of psychological factors in chronic pain. *Acta Anaesthesiologica Scandinavica, 43*(9), 885-888.

Turk, D. & Burwinkle, T. (2005). Assessment of chronic pain in rehabilitation: outcome measures in clinical trials and clinical practice. *Rehabilitation Psychology, 50*(1), 56-64.

Turk, D., Dworkin, R., Burke, L., Gershon, R., Rothman, M., Scott, J., et al. (2006). Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain, 125*(3), 208-215.

Turk, D., Dworkin, R., McDermott, M., Bellamy, N., Burke, L., Chandler, J., et al. (2008). Analyzing multiple endpoints in clinical trials of pain treatments: IMMPACT recommendations. *Pain, 139*(3), 485-493.

Turk, D., Dworkin, R., Revicki, D., Harding, G., Burke, L., Cella, D., et al. (2008). Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. *Pain, 137*(2), 276-285.

Turk, D. & Okifuji, A. (2002). Psychological factors in chronic pain: evolution and revolution. *Journal of Consulting and Clinical Psychology, 70*(3), 678-690.

Turk, D., Swanson, K. & Tunks, E. (2008). Psychological approaches in the treatment of chronic pain patients - when pills, scalpels, and needles are not enough. *Canadian Journal of Psychiatry, 53*(4), 213-223.

University of Bristol (2009). Bristol Online Surveys Retrieved 26th January 2009, from http://www.survey.bris.ac.uk

Van Damme, S., Crombez, G., Goubert, L. & Eccleston, C. (2009). Current issues and new directions in psychology and health: the costs and benefits of self-regulation – a call for experimental research. *Psychology and Health, 24*(4), 367-371.

Van Tulder, M., Assendelft, W., Koes, B. & Bouter, L. (1997). Spinal radiographic findings and nonspecific low back pain: a systematic review of observational studies. *Spine, 22*(14), 427-434.

Vowles, K. & McCracken, L. (2008). Acceptance and values-based action in chronic pain: a study of treatment effectiveness and process. *Journal of Consulting and Clinical Psychology, 76*(3), 397-407.

Wade, W., Treat, T. & Stuart, G. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: a benchmarking strategy. *Journal of Consulting and Clinical Psychology, 66*(2), 231-239.

Wampold, B. (2001). Relative Efficacy: the dodo bird was smarter than we have been led to believe. In B. Wampold (Ed.), *The Great Psychotherapy Debate: Models, Methods and Findings* (pp. 72-118). New Jersey: Lawrence Erlbaum Associates.

Wampold, B., Goodheart, C. & Levant, R. (2007). Clarification and elaboration on evidence-based practice in psychology. *American Psychologist 62*(6), 616-618.

Wampold, B., Mondin, G., Moody, M., Stich, F., Benson, K. & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically, "all must have prizes". *Psychological Bulletin, 122*(3), 203-215.

Waring, P. & Booth, A. (2006). UK Pain Management Programme (PMP) Directory 2006. Liverpool: Walton Centre for Neurology and Neurosurgery.

Weersing, V. & Weisz, J. (2002). Community clinic treatment of depressed youth: benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology, 70*(2), 299-310.

White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 41-55). New York: Russell Sage Foundation

Yates, S., Morley, S., Eccleston, C. & Williams, A. (2005). A scale for rating the quality of psychological trials for pain. *Pain, 117*(3), 314-325.

Zigmond, A. & Snaith, R. (1983). The hospital anxiety and depression scale. *Acta Psyciatrica Scandinavica, 67*(6), 361-370.

REFERENCES TO INCLUDED STUDIES

Note. These are the primary references to included studies: those marked with an asterisk were described in more than one publication.

Altmaier, E., Lehmann, T., Russell, D., Weinstein, J. & Kao, C. (1992). The effectiveness of psychological interventions for the rehabilitation of low back pain: a randomized controlled trial evaluation. *Pain, 49*(3), 329-335.

Astin, J., Berman, B., Bausell, B., Lee, W., Hochberg, M. & Forys, K. (2003). The efficacy of mindfulness meditation plus qigong movement therapy in the treatment of fibromyalgia: a randomized controlled trial. *Journal of Rheumatology 30*(10), 2257-2262.

Basler, H., Jakle, C. & Kroner-Herwig, B. (1997). Incorporation of cognitive-behavioral treatment into the medical care of chronic low back patients: a controlled randomized study in German pain treatment centers. *Patient Education and Counseling 31*(2), 113-124.

Becker, N., Sjogren, P., Bech, P., Olsen, A. & Eriksen, J. (2000). Treatment outcome of chronic non-malignant pain patients managed in a Danish multidisciplinary pain centre compared to general practice: a randomised controlled trial. *Pain, 84*(2-3), 2-3-211.

Bliokas, V., Cartmill, T. & Nagy, B. (2007). Does systematic graded exposure in vivo enhance outcomes in multidisciplinary chronic pain management groups? *Clinical Journal of Pain, 23*(4), 361-374.

Carson, J., Keefe, F., Lynch, T., Carson, K., Goli, V., Fras, A., et al. (2005). Loving-kindness meditation for chronic low back pain: results from a pilot trial. *Journal of Holistic Nursing 23*(3), 287-304.

Cook, A. (1998). Cognitive-behavioral pain management for elderly nursing home residents. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 53B*(1), 51-59.

Ersek, M., Turner, J., Cain, K. & Kemp, C. (2008). Results of a randomized controlled trial to examine the efficacy of a chronic pain self-management group for older adults. *Pain, 138*(1), 29-40.

Ersek, M., Turner, J., McCurry, S., Gibbons, L. & Kraybill, B. (2003). Efficacy of a self-management group intervention for elderly persons with chronic pain. *Clinical Journal of Pain, 19*(3), 156-167.

Evers, A., Kraaimaat, F., Van Riel, P. & De Jong, A. (2002). Tailored cognitive-behavioral therapy in early rheumatoid arthritis for patients at risk: a randomized controlled trial. *Pain, 100*(1-2), 141-153.

Flor, H. & Birbaumer, N. (1993). Comparison of the efficacy of electromyographic biofeedback, cognitive-behavioral therapy, and conservative medical interventions in the treatment of chronic musculoskeletal pain. *Journal of Consulting and Clinical Psychology, 61*(4), 653-658.

Greco, C., Rudy, T. & Manzi, S. (2004). Effects of a stress-reduction program on psychological function, pain, and physical function of systemic lupus erythematosus patients: a randomized controlled trial. *Arthritis and Rheumatism, 51*(4), 625-634.

Haldorsen, E., Kronholm, K., Skouen, J. & Ursin, H. (1998). Multimodal cognitive behavioral treatment of patients sicklisted for musculoskeletal pain: a randomized controlled study. *Scandinavian Journal of Rheumatology, 27*(1), 16-25.

Jensen, I., Bergström, G., Ljungquist, T., Bodin, L. & Nygren, A. (2001). A randomized controlled component analysis of a behavioral medicine rehabilitation program for chronic spinal pain: are the effects dependent on gender? *Pain, 91*(1), 65-78.

Johansson, C., Dahl, J., Jannert, M., Melin, L. & Andersson, G. (1998). Effects of a cognitive-behavioral pain-management program. *Behaviour Research and Therapy, 36*(10), 915-930.

*Keefe, F., Caldwell, D., Baucom, D., Salley, A., Robinson, E., Timmons, K., et al. (1996). Spouse-assisted coping skills training in the management of osteoarthritic knee pain. *Arthritis Care and Research, 9*(4), 279-291.

*Keefe, F., Caldwell, D., Williams, D., Gil, K., Mitchell, D., Robertson, C., et al. (1990). Pain coping skills training in the management of osteoarthritic knee pain: a comparative study. *Behavior Therapy, 21*(1), 49-62.

Kraaimaat, F., Brons, M., Geenen, R. & Bijlsma, J. (1995). The effect of cognitive behavior therapy in patients with rheumatoid arthritis. *Behaviour Research and Therapy, 33*(5), 487-495.

Linton, S., Boersma, K., Jansson, M., Overmeer, T., Lindblom, K. & Vlaeyen, J. (2008). A randomized controlled trial of exposure in vivo for patients with spinal pain reporting fear of work-related activities. *European Journal of Pain, 12*(6), 722-730.

Mishra, K., Gatchel, R. & Gardea, M. (2000). The relative efficacy of three cognitive-behavioral treatment approaches to temporomandibular disorders. *Journal of Behavioral Medicine, 22*(3), 293-309.

Moore, J. & Chaney, E. (1985). Outpatient group treatment of chronic pain: effects of spouse involvement. *Journal of Consulting and Clinical Psychology, 53*(3), 326-334.

Morone, N., Greco, C. & Weiner, D. (2008). Mindfulness meditation for the treatment of chronic low back pain in older adults: a randomized controlled pilot study. *Pain, 134*(3), 310-319.

Newton-John, T., Spence, S. & Schotte, D. (1995). Cognitive-behavioral therapy versus EMG biofeedback in the treatment of chronic low back pain. *Behaviour Research and Therapy, 33*(6), 691-697.

Nicassio, P., Radojevic, V., Weisman, M., Schuman, C., Kim, J., Schoenfeld-Smith, K., et al. (1997). A comparison of behavioral and educational interventions for fibromyalgia. *Journal of Rheumatology, 24*(10), 2000-2007.

Puder, R. (1988). Age analysis of cognitive-behavioral group therapy for chronic pain outpatients. *Psychology and Aging, 3*(2), 204-207.

Radojevic, V., Nicassio, P. & Weisman, M. (1992). Behavioral intervention with and without family support for rheumatoid arthritis. *Behavior Therapy, 23*(1), 13-30.

Thieme, K., Gromnica-Ihle, E. & Flor, H. (2003). Operant behavioral treatment of fibromyalgia: a controlled study. *Arthritis and Rheumatism 49*(3), 314-320.

Turner, J. & Clancy, S. (1988). Comparison of operant behavioral and cognitive-behavioral group treatment for chronic low back pain. *Journal of Consulting and Clinical Psychology, 56*(2), 261-266.

Turner, J., Clancy, S., McQuade, K. & Cardenas, D. (1990). Effectiveness of behavioral therapy for chronic low back pain: a component analysis. *Journal of Consulting and Clinical Psychology, 58*(5), 573-579.

Turner, J. & Jensen, M. (1993). Efficacy of cognitive therapy for chronic low back pain. *Pain, 52*(2), 169-177.

Turner, J. Mancl, L. & Aaron, L. (2006). Short- and long-term efficacy of brief cognitive-behavioral therapy for patients with chronic temporomandibular disorder pain: a randomized, controlled trial. *Pain, 121*(3), 181-194.

Vlaeyen, J., Haazen, I., Schuerman, J., Kole-Snijders, A. & Van Eek, H. (1995). Behavioural rehabilitation of chronic low back pain: comparison of an operant treatment, an operant-cognitive treatment and an operant-respondent treatment. *British Journal of Clinical Psychology, 34*(1), 95-118.

Vlaeyen, J., Teeken-Gruben, N., Goossens, M., Rutten-Van Molken, M., Pelt, R., Van Eek, H., et al. (1996). Cognitive-educational treatment of fibromyalgia: a randomized clinical trial. I: clinical effects. *Journal of Rheumatology, 23*(7), 1237-1245.

Williams, A., Richardson, P., Nicholas, M., Pither, C., Harding, V., Ridout, K., et al. (1996). Inpatient vs. outpatient pain management: results of a randomised controlled trial. *Pain, 66*(1), 13-22.

Woods, M. & Asmundson, G. (2008). Evaluating the efficacy of graded *in vivo* exposure treatment for the treatment of fear in patients with chronic back pain: a randomized controlled clinical trial. *Pain, 136*(3), 271-280.

Zautra, A., Davis, M., Reich, J., Nicassio, P., Tennen, H., Finan, P., et al. (2008). Comparison of cognitive behavioral and mindfulness meditation interventions on adaptation to rheumatoid arthritis for patients with and without history of chronic depression. *Journal of Consulting and Clinical Psychology, 76*(3), 408-421.

REFERENCES TO EXCLUDED STUDIES

Note. These are the primary references to excluded studies: those marked with an asterisk were described in more than one publication.

Babu, A., Mathew, E., Danda, D. & Prakash, H. (2007). Management of patients with fibromyalgia using biofeedback: a randomized controlled trial. *Indian Journal of Medical Science, 61*(8), 455-461.

Bradley, L., Young, L., Anderson, K., Turner, R., Agudelo, C., McDaniel, L., et al. (1987). Effects of psychological therapy on pain behavior of rheumatoid arthritis patients. Treatment outcome and six-month follow-up. *Arthritis and Rheumatism, 30*(10), 1105-1114.

Buckelew, S., Conway, R., Parker, J., Deuser, W., Read, J., Witty, T., et al. (1998). Biofeedback/relaxation training and exercise interventions for fibromyalgia: a prospective trial. *Arthritis Care and Research 11*(3), 196-209.

Buhrman, M., Faltenhag, S., Strom, L. & Andersson, G. (2004). Controlled trial of Internet-based treatment with telephone support for chronic back pain. *Pain, 111*(3), 368-377.

Dworkin, S., Huggins, K., Wilson, L., Mancl, L., Turner, J., Massoth, D., et al. (2002). A randomized clinical trial using research diagnostic criteria for temporomandibular disorders-axis II to target clinic cases for a tailored self-care TMD treatment program. *Journal of Orofacial Pain, 16*(1), 48-63.

Dworkin, S., Turner, J., Wilson, L., Massoth, D., Whitney, C., Huggins, K., et al. (1994). Brief group cognitive-behavioral intervention for temporomandibular disorders. *Pain, 59*(2), 175-187.

*Fairbank, J., Frost, H., Wilson-MacDonald, J., Yu, L., Barker, K. & Collins, R. (2005). Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *British Medical Journal, 330*(1), 1-7.

Freeman, K., Hammond, A. & Lincoln, N. (2002). Use of cognitive-behavioural arthritis education programmes in newly diagnosed rheumatoid arthritis. *Clinical Rehabilitation 16*(8), 828-836.

* Geraets, J., Goossens, M., De Bruijn, C., Koke, A., De Bie, R., Pelt, R., et al. (2004). A behavioural treatment for chronic shoulder complaints: concepts, development, and study design. Australian Journal of Physiotherapy, 50(1), 33-38.

*Hammond, A. & Freeman, K. (2001). One-year outcomes of a randomized controlled trial of an educational-behavioural joint protection programme for people with rheumatoid arthritis. *Rheumatology 40*(9), 1044-1051.

Jensen, I., Bergström, G., Ljungquist, T. & Bodin, L. (2005). A 3-year follow-up of a multidisciplinary rehabilitation programme for back and neck pain. *Pain, 115*(3), 273-283.

Jensen, I., Dahlquist, C., Nygren, A., Royen, E. & Sternberg, M. (1997). Treatment for 'helpless' women suffering from chronic spinal pain: a randomized controlled 18-month follow-up study. *Journal of Occupational Rehabilitation, 7*(4), 225-238.

Kaapa, E., Frantsi, K., Sarna, S. & Malmivaara, A. (2006). Multidisciplinary group rehabilitation versus individual physiotherapy for chronic nonspecific low back pain: a randomized trial. *Spine, 31*(4), 371-376.

*Kole-Snijders, A., Vlaeyen, J., Goossens, M., Van Moelken, R., Heuts, P., Van Breukelen, G., et al. (1999). Chronic low-back pain: what does cognitive coping skills training add to operant behavioral treatment? Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 67*(6), 931-944.

Leeuw, M., Goossens, M., Van Breukelen, G., De Jong, J., Heuts, P., Smeets, R., et al. (2008). Exposure in vivo versus operant graded activity in chronic low back pain patients: results of a randomized controlled trial. *Pain, 138*(1), 192-207.

Lorig, K., Ritter, P., Laurent, D. & Plant, K. (2008). The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. *Arthritis Care and Research 59*(7), 1009-1017.

Marhold, C., Linton, S. & Melin, L. (2001). A cognitive-behavioral return-to-work program: effects on pain patients with a history of long-term versus short-term sick leave. *Pain, 91*(1-2), 155-163.

McCarberg, B. & Wolf, J. (1999). Chronic pain management in a health maintenance organization. *The Clinical Journal of Pain, 15*(1), 50-57.

O'Leary, A., Shoor, S., Lorig, K. & Holman, H. (1988). A cognitive-behavioral treatment for rheumatoid arthritis. *Health Psychology, 7*(6), 527-544.

Parker, J., Frank, R., Beck, N., Smarr, K., Buesher, K., Phillips, L., et al. (1988). Pain management in rheumatoid arthritis patients. A cognitive behavioural approach. *Arthritis and Rheumatism 31*(5), 593-601.

Redondo, J., Justo, C., Moraleda, F., Velayos, Y., Puche, J., Zubero, J., et al. (2004). Long-term efficacy of therapy in patients with fibromyalgia: a physical exercise-based program and a cognitive-behavioral approach. *Arthritis and Rheumatism, 51*(2), 184-192.

*Smeets, R., Vlaeyen, J., Hidding, A., Kester, A., Van Der Heijden, G., Van Geel, C., et al. (2006). Active rehabilitation for chronic low back pain: Cognitive-behavioral, physical, or both? First direct post-treatment results from a randomized controlled trial. *BMC Musculoskeletal Disorders, 7*(5), 1-16.

*Spence, S. (1989). Cognitive-behavior therapy in the management of chronic, occupational pain of the upper limbs. *Behavior Research and Therapy, 27*(4), 435-446.

Strauss, G., Spiegel, J., Daniels, M., Speigel, T., Landsverk, J., Roy-Byne, P., et al. (1986). Group therapies for rheumatoid arthritis: a controlled study of two approaches. *Arthritis and Rheumatism 29*(10), 1203-1209.

Strong, J. (1998). Incorporating cognitive-behavioral therapy with occupational therapy: a comparative study with patients with low back pain. *Journal of Occupational Rehabilitation, 8*(1), 61-71.

Turner-Stokes, L., Erkeller-Yuksel, F., Miles, A., Pincus, T., Shipley, M. & Pearce, S. (2003). Outpatient cognitive behavioral pain management programs: a randomized comparison of a group-based multidisciplinary versus an individual therapy model. *Archives of Physical Medicine and Rehabilitation 84*(6), 781-788.

APPENDICES

Appendix 1: Correspondence Relating to Ethical Approval

Date: Mon, 16 Feb 2009 13:47:15 +0000 [16/02/2009 13:47:15 GMT]
From: Amy Beckitt <Amy.Beckitt@leedsth.nhs.uk>
To: Grania Fenton <ugm6g2f@leeds.ac.uk>
Subject: Re: ethics confirmation for PMP survey

Hi Grania,

Further to our telephone conversation and email correspondence, please note that the below proposal has been reviewed by a Chair of an NHS Research Ethics Committee and resultantly, the proposal does not require NHS REC approval.

Should you require any further information, please do not hesitate to ask.

Kind Regards,

Amy Beckitt
Committee Assistant Coordinator
Leeds (East) Research Ethics Committee
Room 5.2
Clinical Sciences Building
St James's University Hospital
Leeds
LS9 7TF

Tel: 0113 206 5637
Fax: 0113 206 6772
Email: amy.beckitt@leedsth.nhs.uk

>>> Grania Fenton <ugm6g2f@leeds.ac.uk> 16/02/2009 13:28 >>>
Hi Amy,

the proposal is as follows:

To ask lead NHS staff from UK Pain Management Programmes (PMPs) to complete a brief online survey about broad characteristics of their client group and the services offered to them. This is to ascertain how comparable routine clinical activity is to descriptions of treatments and participants within published RCTs, and therefore ultimately how applicable RCT data is to clinical settings.

Participation will be entirely voluntary, and the survey will be advertised at the Pain Society conference, in a newsletter and possibly through an email sent to all PMP leads. At the end of the

survey participants may be asked to email me if they would like to participate in a further brief electronic study to gain a consensus on outcome measures that are useful within PMPs. The aim of this brief study is to ensure that the ways in which I interrogate the published RCT data with regards outcome measures is as useful to PMPs as possible. Let me know if you've any further questions,

Kind regards,

Grania Fenton

Appendix 2: Survey Questions (study 1)

This PMP survey has four sections regarding team, patient, treatment and outcome characteristics.

If you would prefer not to answer a particular question please select don't know/prefer not to say rather than leaving it blank.

Once you have completed all four sections, please click on the continue button at the bottom of the page. Please note that once you have done this, your responses will be automatically submitted and you will not be able to return to review or amend them.

Team characteristics

1. Please state the PMP that you work for in the below space (n.b. this information will be treated with the strictest confidence and simply help to ensure a representative sample).

a. Please state whether you are responding to this survey on behalf of your team or as an individual

How long have you worked here?

< 6 months, 6 months to a year, 1 - 3 years, 3 - 5 years, > 5 years, Don't know/prefer not to say

b. Are you a member of the British Pain Society?

Yes, No, Don't know/prefer not to say

c. Please select your core profession from the list below

Administrative staff, Assistant psychologist, CBT therapist, Clinical psychologist, Counselling psychologist, Graduate patient, Health psychologist, Medical, Nurse
Occupational therapist, Pharmacist, Physiotherapist, Psychologist in Clinical Training, Don't know/prefer not to say, Other (please specify):

d. Please select all of the professions who work within your PMP

(select all that apply)

Administrative staff, Assistant psychologist, CBT therapist, Clinical psychologist, Counselling psychologist, Graduate patient, Health psychologist, Medical, Nurse
Occupational therapist, Pharmacist, Physiotherapist, Psychologist in Clinical Training, Don't know/prefer not to say, Other (please specify):

2. Does your PMP routinely provide an induction period for new staff?
Yes, No, Don't know / prefer not to say

3. Does your PMP routinely provide systematic clinical supervision for clinical staff?
Yes, No, Don't know / prefer not to say

4. Does your PMP routinely provide therapeutic training for clinicians (specifically regarding patients with chronic pain) before they start working with patients?
Yes, No, Don't know / prefer not to say

5. Would you describe your PMP team as 'inter-disciplinary' (i.e. one in which diverse professionals work together on a routine basis)?
Yes, No, Don't know / prefer not to say

6. Would you agree that staff members are encouraged to allocate time for continuing professional development on a regular basis?
Yes, No, Don't know / prefer not to say

7. If you have any further comments about these team characteristics questions, please write them in the space provided below

Patient characteristics

I appreciate that some Pain Management Programmes may not collect some types of patient data routinely or that it may not be easily accessible, so please simply provide your best estimate to questions in this section.

8. Approximately how many patients are seen within your PMP in a given year? and approximately what percentage of these patients are female?

9. What would you estimate the average age of patients to be?
and what would you estimate the approximate age range of patients to be? (E.g. 18-60 etc.)

10. What percentage of patients would you estimate to be currently employed (either part-time or full time)?

11. Please estimate the approximate percentage of patients that may fall into each diagnostic category. (I appreciate that this is an estimate and so the percentage may not add up to 100

Don't know/prefer not to say

0%, <5%, 5-10%, 11-20%, 21-30%, 31-40%, 41-50%, 51-60%, 61-70%, 71-80%, 81-90%, 91-100%

a. Chronic low back pain b. Complex regional pain syndrome c. Facial

d. Fibromyalgia e. Headache f. Multiple sites of pain g. Musculoskeletal

h. Neuralgia i. Osteo-arthritis j. Rheumatoid arthritis k. Temporomandibular joint syndrome l. Upper limb m. Visceral n. Other

12. If you have selected 'other' in the diagnostic category question above, please specify the diagnosis/diagnoses and approximate percentage of patients in the space below.

13. If you have any further comments about these patient characteristics questions, please write them in the space provided below.

Treatment characteristics

14. What would you consider to be the main treatment orientation of this PMP?

Acceptance and Commitment Therapy (i.e. a psychological intervention which incorporates acceptance, mindfulness and value oriented behavioural strategies).

Behaviour Therapy (i.e. a psychological intervention based on principles of operant conditioning and social reinforcement, involving techniques such as identifying and reducing unhelpful patterns of behaviour and building upon helpful patterns of behaviour in order to facilitate change).

Biofeedback (i.e. techniques which use external feedback to raise levels of conscious control over body functions such as muscle tension, with the aim of improving control over such functions. May also include relaxation components but NOT as the main focus of therapy).

Cognitive Behaviour Therapy (i.e. a goal oriented psychological therapy involving examination of the links between cognitions, emotions, physiology and behaviour. Involves the use of Socratic questioning and collaborative empiricism e.g. testing beliefs and assumptions through behavioural experiments, in addition to techniques such as activity scheduling, in order to influence change).

Mindfulness (i.e. teaching skills in paying attention in the present moment, on purpose and non-judgementally using specific 'mindfulness meditations').

Mixed (i.e. a combination of several of the treatments mentioned in this section. If you select this option, please specify the treatments in the space below).

Relaxation (i.e. the use of a specific training strategy designed to induce physical or mental relaxation).

Don't know/prefer not to say

Other (please specify)

a. Please indicate the average length of this treatment in number of weeks (excluding any follow-up/booster sessions)

b. Please indicate the average number of hours per week of this treatment (excluding any follow-up/booster sessions)

c. Do you routinely offer follow-up/booster sessions?
Yes, No, Don't know / prefer not to say

d. If yes, please state the average length of time in weeks between the start of treatment and the first follow-up/booster session(s)

e. Do you routinely follow-up patients who drop out of treatment?
Yes, No, Don't know / prefer not to say

f. If yes, please use the below space to detail how you follow-up patients who drop out and what you do with the information

15. In consideration of the main treatment orientation as specified above, please indicate the individual components of treatment

a. Is there a specific education component within the main treatment specified above?

Yes, No, Don't know / prefer not to say

b. Is there a specific exercise component within the main treatment specified above?

Yes -- managed by non-psychological principles (e.g. physiotherapy)

Yes -- managed by psychological principles (e.g. graded exposure)

No, Don't know / prefer not to say

c. Is there a specific goal setting component within the main treatment specified above?

Yes, No, Don't know / prefer not to say

d. Is there a specific problem solving component within the main treatment specified above?

Yes -- it is the main focus of treatment

Yes -- but it is not the main focus of the treatment

No, Don't know / prefer not to say

e. Is there a specific relaxation component within the main treatment specified above? (i.e. the use of a specific training strategy designed to induce physical or mental relaxation, e.g. progressive muscle relaxation, relaxation with imagery. Please record biofeedback or mindfulness/meditation below).

Yes -- some strategies are taught within the main treatment

Yes -- it is the main focus of the treatment

No, Don't know/prefer not to say

f. If yes, please state the type of relaxation that is a component of treatment in the space below

g. Is there a specific biofeedback component within the main treatment specified above? (i.e. a component which raises patients' awareness and levels of conscious control over body functions such as muscle tension).
Yes -- some strategies are taught within the main treatment
Yes -- it is the main focus of the treatment
No, Don't know / prefer not to say

h. Is there a specific behavioural management component within the main treatment specified above? (i.e. using concepts of operant conditioning etc. as described in behaviour therapy above, but not as the main focus of treatment).
Yes -- some strategies are taught within the main treatment
Yes -- it is the main focus of the treatment
No, Don't know / prefer not to say

i. Is there a specific attention management component within the main treatment specified above? (i.e. teaching attention management skills with the aim of limiting the impact of pain and enhancing the sense of control over it, either by switching attention to another stimulus complex (internal or external) or by 'retuning' to the pain stimulus with the aim of altering the appraisal of it. Includes attention-diversion training, distraction, and sensory reinterpretation. Please record mindfulness/meditation below).
Yes - some strategies are taught within the main treatment
Yes - it is the main focus of the treatment
No, Don't know / prefer not to say

j. Is there a specific mindfulness/meditation component within the main treatment specified above? (i.e. teaching skills in paying attention in the present moment, on purpose and non-judgementally using specific 'mindfulness meditations'. Please record attention-diversion training, distraction, and sensory reinterpretation above).
Yes - some strategies are taught within the main treatment
Yes - it is the main focus of the treatment
No, Don't know / prefer not to say

k. Is there a specific coping skills component within the main treatment specified above? (i.e. teaching a range of methods to help patients to manage specific pain episodes)

Yes - some strategies are taught within the main treatment
Yes - it is the main focus of the treatment
No, Don't know / prefer not to say

l. Is there a specific cognitive restructuring component within the main treatment specified above? (i.e. aiming to move patients beyond the specifics of coping with a particular pain episode by encouraging them to restructure their appraisals and attributions about their pain, thus changing their relationship towards it).
Yes - some strategies are taught within the main treatment
Yes - it is the main focus of the treatment
No, Don't know / prefer not to say

m. In consideration of the main treatment orientation as specified above, please use the below space to record any other components of treatment

16. Please indicate the setting in which the majority of your PMP's treatments are provided
In-patient
Out-patient community
Out-patient hospital
Don't know/prefer not to say
Other (please specify):

17. Please select the predominant mode of treatment delivery in your PMP (n.b significant others are defined as spouse, partners, family members etc.).
Group
Individual
Mixed group and individual
Mixed group and significant others
Mixed individual and significant others
Mixed group and individual and significant others
Don't know/prefer not to say
Other (please specify):
If significant others are involved in treatment, please select their level of involvement

Significant others are considered to be a part of the treatment (i.e. they attend for one or two specific sessions)
Significant others are considered to be integral to the treatment (i.e. they are involved in a significant proportion of the treatment)
Don't know/prefer not to say

18. Do you use manuals for any of the treatments that you offer within this Pain Management Programme?
Yes -- for some treatments/components of treatments
Yes -- for all treatments/components of treatments
No, Don't know/prefer not to say
If you have indicated yes to the above question, please record the names of the manuals that you use for specific components in the below space

19. If you have any further comments about these treatment characteristics questions, please write them in the space provided below

Outcome characteristics

20. With the main treatment orientation as specified above in mind, please indicate when, if at all, you use outcome measures

	Yes	No	Don't know / prefer not to say
a. before treatment			
b. after treatment			
c. at follow-up			

21. You will be aware that there a number of domains on which outcomes of treatments provided within PMPs can be measured. In consideration of the following domains, please indicate those outcomes that are routinely measured within your PMP.
a. Coping
(select all that apply)
Coping Strategies Questionnaire
Don't know / prefer not to say
Not applicable - don't measure
Other (please specify):

b. Mood

(select all that apply)

Arthritis Impact Measurement Scale (AIMS) Mental Disability

Arthritis Impact Measurement Scale (AIMS) Psychological Disability

Hospital Anxiety and Depression Scale (HADS)

Beck Anxiety Inventory (BAI)

Beck Depression Inventory (BDI)

Brief Symptom Inventory (BSI) Distress

Centre for Epidemiologic Studies Depression Scale (CES-DS)

Clinical Outcomes in Routine Evaluation (CORE)

Depression Adjective Checklist (DACL)

Depression Scale (DEPS)

Geriatric Depression Scale (GDS)

Hopkins Symptom Checklist Distress (HSCD)

Impact of Rheumatic diseases on General health and Lifestyle (IRGL) Depression subscale

Minnesota Multiphasic Personality Inventory (MMPI) Depression subscale

State-Trait Anxiety Inventory State (STAI-S) Subscale

State-Trait Anxiety Inventory Trait (STAI-T) Subscale

(West Haven) Multidimensional Pain Inventory (WHYMPI) Distress subscale Don't know / prefer not to say

Not applicable - don't measure

Other (please specify):

c. Disability

(select all that apply)

Arthritis Impact Measurement Scale (AIMS) Activities of Daily Living

Arthritis Impact Measurement Scale (AIMS) Physical Disability

Disability Rating Index

Dusseldorf Disability Scale

Fibromyalgia Impact Questionnaire

Impact of Rheumatic diseases on General health and Lifestyle (IRGL) Functional Disability

Oswestry Disability Index

Pain Disability Inventory

Roland and Morris Disability Questionnaire

Sickness Impact Profile (SIP; other)

Sickness Impact Profile (SIP; self)

SF-36 Physical Function

(West Haven) Multidimensional Pain Inventory (WHYMPI) Activity subscale

(West Haven) Multidimensional Pain Inventory (WHYMPI) Pain Interference subscale

Don't know / prefer not to say

Not applicable - don't measure

Other (please specify):

d. Pain behaviour

Direct observation

Exercise tests (e.g. five minute walk, sit to stand)

Medication count

Self-report schedule (please specify which one below if so)

Don't know / prefer not to say

Not applicable - don't measure

Other (please specify):

e. Pain experience

(select all that apply)

Arthritis Impact Measurement Scale (AIMS)

Arthritis Self-Efficacy Scale

Brief Pain Inventory (BPI)

Characteristic Pain Intensity

Graded Chronic Pain Scale

Impact of Rheumatic diseases on General health and Lifestyle (IRGL)

McGill Pain Questionnaire Pain Rating Index

Memorial Pain Questionnaire

Multi-dimensional Pain Inventory (MPI) Pain Severity Rating

Numerical Rating Scale

SF-36 Bodily Pain Scale

Verbal Rating Scale

Visual Analogue Scale

Don't know / prefer not to say

Not applicable - don't measure

Other (please specify):

f. Social role functioning (i.e. how much pain interferes with social roles)

(select all that apply)

Pain Disability Inventory

Sickness Impact Profile (SIP; other)

Sickness Impact Profile (SIP; self)

Visual Analogue Scale of Interference to social role functioning

(West Haven) Multi-dimensional Pain Inventory (MPI) Pain Interference subscale

Don't know / prefer not to say

Not applicable - don't measure

Other (please specify):

g. Health and social care use by the patient

(select all that apply)

Employment status

Number of visits to GP

Number of visits to other medical settings

Medication Use

Don't know / prefer not to say

Not applicable - don't measure

22. Have the outcomes of treatments within this PMP ever been written up for clinical audit, research or service evaluation purposes?

Yes, No, Don't know / prefer not to say

If yes, please detail any publications

23. If you have any further comments about these outcome characteristics questions or would like to record any other outcome measures that your PMP administers, please use the below space.

Appendix 3: Preliminary Survey Results (study 1)

All seventy-six members of the British Pain Society's Pain Management Programme Special Interest Group[1] (PMP SIG) and attendees of the most recent PMP conference were invited to participate in this survey between May and August 2009.

Thirty-six separate responses were received in total, all of which were from respondents involved with services in England. Thirty-three responses appeared to be from people within different NHS based PMPs, several from within the same Trust. One respondent was from an Expert Patient Programme and one was from an independent healthcare provider working with the NHS. Additionally, one respondent chose not to state the PMP in which they worked, but I have chosen to summarise data from all respondents at this stage of the analysis.

One third of responses (n=12) were on behalf of PMP teams, with the remaining two-thirds classified as individuals' responses.

The main results were as follows[2]:

Team Characteristics

- 74% of respondents (n=26) were members of the British Pain Society.
- The majority of respondents (61%; n=22) were physiotherapists or clinical psychologists. The remaining classified themselves as belonging to the medical, nursing, occupational therapy or clinical and health psychology professions, with one respondent describing themselves as a trainer.
- 92% of respondents (n=33) classed themselves as belonging to an 'inter-disciplinary' team where multiple professions worked together on a routine basis. Teams were frequently made up of administrative staff, clinical psychologists, physiotherapists and members of the medical and nursing professions, although several respondents also stated that assistant and trainee clinical psychologists, CBT therapists, counselling and

[1] British Pain Society (2009) Pain Management Programme Society SIG. Retrieved 24th August 2009, from http://www.britishpainsociety.org/members_sigs_pmp.htm

[2] Please note that questions completion rates were not 100% and so the total n used to calculate overall percentages is varied. Percentages reported are approximate and so totals may not add up to 100.

health psychologists, graduate patients, pharmacists and occupational therapists made contributions to the team.

- With regards professional development, two-thirds of respondents (n=24) stated that an induction programme was routine for new staff and 61% (n=22) stated that systematic clinical supervision was routinely provided. Whilst only 36% (n=13) stated that therapeutic training was routinely provided for clinicians, 86% (n=31) agreed that staff were encouraged to allocate time for continuing professional development on a regular basis.
- Several respondents suggested that the manner in which continuing professional development, clinical supervision, inter-disciplinary working and induction periods were encouraged and provided was variable.

Patient Characteristics

- Estimates of the number of patients seen per annum varied widely between 20 and 3000, reflecting the varied capacities and clinical foci of the services described by respondents. However, the vast majority (61%; n=22) estimated that fewer than 100 patients were treated within PMPs per annum, 25% (n=9) estimated between 100 and 200, and 8% (n=3) between 200 and 400. Only 4% (n=2) of respondents, both of whom worked within NHS PMPs, estimated that more than 1500 patients were treated annually.
- 86% of respondents (n=30) estimated that at least 60% of patients treated were female.
- Although all respondents estimated that the average age of patients was between 40 and 50, services did not appear to have imposed upper age limits. However, only 5% of respondents (n=2) implied that adolescent referrals would be accepted within their service.
- The vast majority of respondents (81%; n=29) estimated that fewer than one third of patients treated were in any form of employment, and only one respondent suggested that this was due to many patients being outside the usual working age range.
- It was recognised that many patients have multiple sites of pain, and many respondents reported that service inclusion criteria influenced the likely range of presenting diagnoses and sites of pain reported by their patients. However, respondents estimated the following: 46% (n=16) that at least 70% of patients had

musculoskeletal pain; 58% (n=21) that at least 61% had chronic low back pain; 69% (n=25) that at least 11% had fibromyalgia; 47% (n=17) that at least 11% had osteoarthritis and 40% (n=14) that at least 11% had upper limb pain. Fewer patients were estimated to be likely to present with complex regional pain syndrome, facial pain, headache, inflammatory arthropathy such as rheumatoid arthritis or lupus, temporomandibular joint syndrome, vascular, post-operative or visceral pain.

Treatment Characteristics

- Only one respondent chose not to report the main treatment orientation of their programme. The vast majority cited Cognitive Behavioural Therapy (CBT) as the main orientation (42%; n=15), with a further 17% (n=6) citing CBT in addition to components of other treatments.
- 5% of respondents (n=2) described Acceptance and Commitment Therapy (ACT) as the main theoretical orientation of their programme, with the same proportion stating that ACT was a component of the main treatment. One respondent described a recent service move to ACT from CBT, and mindfulness methods were described as being employed where pain was not related to activity (n=1) or where many patients did not have English as a first language (n=1).
- 73% of respondents (n=24) reported that their programme ran for at least twenty-five hours, and although the average number of treatment hours was forty-nine, this ranged from three hours to over one hundred.
- 53% of respondents (n=18) reported a treatment duration of at least eight weeks, and although the average duration was 6.8 weeks, this ranged from three to thirteen weeks in total. One respondent specifically described a two-week long residential programme, although 18% of respondents (n=6) implied that their programme comprised similarly intensive non-residential programmes.
- 86% (n=30) of respondents stated that treatments were conducted in out-patient settings; two-thirds of which were in hospital and one-third of which were in community locations.
- 89% of respondents (n=31) stated that their programme offered follow-ups to patients, and 40% (n=14) that they routinely followed-up patients who dropped out of treatment; usually by telephone or in writing.

- The vast majority of respondents (more than 90%) described treatments as containing an educational component, an exercise programme, teaching goal-setting and coping skills or some methods of relaxation. Relaxation components frequently comprised autogenic, progressive muscular, imagery or mindfulness exercises. More than 70% of respondents described problem-solving, specific mindfulness or cognitive restructuring components of treatment and 68% (n=23) described behavioural management components. Biofeedback was reported as a component of treatment by 35% of respondents (n=12).
- 58% of respondents (n=21) reported that treatments were delivered in groups, 28% (n=10) described mixed methods of delivery, and only 6% (n=2) reported any level of individual treatment.
- Of the twenty respondents who answered a question about the involvement of significant others in the programme, 85% (n=17) implied their involvement in several specific treatment sessions but only one respondent described them as being integral to the whole treatment package.
- 51% of respondents (n=18) reported that manuals were used within the treatments offered, and half of these respondents implied that manuals guided all aspects of treatment. Many respondents also reported that handouts were an integral part of the treatment, and all manuals described by respondents had been developed in-house.

Outcome Characteristics

- The vast majority of respondents (97%; n=34) described administering outcome measures before treatment, with slightly fewer using them post-treatment (89%; n=31) and at follow-up (86%; n=30).
- All respondent reported measures of coping, mood, disability and pain experience were standardised.
- Outcome measures within the domains of coping, mood, disability and health and social care use were most likely to be administered within PMPs. There was some consistency within the domains of coping (the Coping Strategies Questionnaire or Pain Self-Efficacy Questionnaire were commonly used), mood (the Hospital Anxiety and Depression Scale was the most commonly employed measure), and health and social care use (employment status, medication use and number of GP and other medical

service visits were commonly recorded). However, there was less consistency with regards the measures of disability that were administered.

- Several respondents stated that their PMP did not routinely administer measures within the following domains: social role functioning (40%; n=14); coping (25%; n=9); pain behaviour (20%; n=7); mood (8%; n=3) or disability (8%; n=3).
- 63% of respondents (n=22) stated that treatment outcomes had been written up, typically for internal auditing purposes. However, several respondents reported that outcome data had comprised post-graduate research, conference posters or been published.

A more formal comparison of the data collected within this survey and the published clinical trial data is currently in progress and will be completed by the end of 2010, but in the meantime if you have any queries about these findings, please contact:

Grania Fenton (contact details supplied)

Appendix 4: Modified Data Extraction Form

General Information

Date of data extraction: _____
Study Identifiers

Study/Article Number: _____

Author(s): _____
Title: _____
Journal: _____Year: _____Vol/Pages: _____
Type of Report:

1. Full report ☐ 2. Follow-up data only ☐ If 2 give study Number of original _____

Location of study:

1. Europe ☐ 2. North America ☐ 3. Australia & NZ ☐ 4. Other (state)_____

Participants

Source of sample:		Sampling strategy:	
Pain or rehab clinic	☐	1. Random from specified pop.	☐
Community referral	☐	2. Convenience from specified pop.	☐
Volunteer	☐	3. Consecutive referrals	☐
Other _____	☐	4. Not reported	☐

Total sample size before selection or attrition: _____
Number entered into trial, after selection / refusal: _____
Number of participants at end of treatment: _____
Number of participants at end of final follow-up: _____
Number of males at start of trial: _____
Number of females at start of trial: _____
Mean age of total sample (to 2 d.p)** _____: ____
SD of sample age: _____: ____
Lowest age: _____
Highest age: _____
*** If age not reported: note means, SD and N for each group page 2.
Averaged years in education _____ If not reported note the page where details of education given
Work status of sample – give as much detail as possible:

1. % employed _____
2. % not employed because of pain _____
3 % not employed for any other reason _____

Mean years of pain of sample: ____:____ or Median if reported _____

Number of previous non-surgical treatments: ___:___
Number of previous surgeries: ____:___Percent of sample with previous surgery:

Diagnostic label for sample:

Chronic low back pain ☐
Rheumatoid arthritis ☐
Osteo-arthritis ☐
Upper limb ☐
Fibromyalgia ☐
TempMandibJoint ☐
Facial ☐
Visceral _____ ☐
CRPS ☐
Multiple sites of pain ☐
Mixed ☐
Other – specify _____

Primary site of pain:

1. Low back ☐
2. Limb ☐
3. Mixed ☐
4. Generalised pain ☐
5. Other – description of sample _____

Notes of on any other aspect of Sample that may need to be discussed

Treatments *** Use separate sheet for each treatment in the trial including control group(s)***
Type and delivery details

Name of treatment as given by author: _____
Treatment Type:
CBT – multiple components ☐
Behaviour therapy ☐
Cognitive therapy ☐
Unimodal treatment e.g. biofeedback ☐ name here _____
Combined e.g. Biofeedback + relaxation ☐ name here _____
Waiting list control – no treatment ☐ i.e. no treatment given by health care
Waiting list -Treatment as usual available ☐ i.e. assigned as WL but TAU available
Treatment as usual control ☐ i.e. assigned to TAU
Education/bibliotherapy control ☐
Other _____

Treatment setting:
In-patient ☐
2. Out-patient Community ☐
3. Outpatient: Hospital ☐
4. Mixed ☐
_____

Format of delivery – patients:
Group ☐
Individual ☐
Group & individual ☐
Individual & spouse ☐
5. Group & spouse ☐
6. Group & individual & spouse ☐
7. Unclear – give comment ☐ _____

Personnel involved:
Psychological expertise

Lay counsellor or trained patient ☐ 4. Clin. Psych qualified +5yrs ☐
Clin. Psych. in training (nurse/CBT therapist) ☐ 5. Qualified non-psychologist ☐
Recently qualified Clin. Psych ☐ 6. Not given ☐

Health care team
1. No details given ☐
2. Multi-professional group of staff involved ☐ i.e. no 'team work'
3. Multidisciplinary team ☐ i.e. definite mention of team

Duration of intervention in weeks: _____Total hours for this treatment: _____
Length of time to follow-up from start of treatment to longest follow up (in weeks): _____

Treatment components

Note If not details are given rate as 0 – not present or if you think that it is present rate it at the lowest level you think appropriate

Education:		Comments here
No education mentioned as part of treatment	☐	
Education mentioned	☐	
Comprehensive education – manualised program	☐	

Exercise:		Comments here
No exercise programme	☐	
Specific programme of exercise mentioned	☐	
Exercise managed by psychological principles	☐	

Goal setting:		
Not explicitly mentioned	☐	Comments here
General statement that used	☐	

Problem Solving:		Comments here
No training given	☐	
Strategies taught in therapy	☐	
PS is the main focus of treatment	☐	

Relaxation:		Comments here
No relaxation	☐	
Relaxation included in treatment	☐	
Relaxation is main focus of treatment	☐	

Biofeedback:		Comments here
No biofeedback	☐	
Biofeedback included in treatment	☐	
Biofeedback is main focus of treatment	☐	

Behavioural management: Comments here
Not present ☐
Behavioural management included in Tx ☐
Behavioural management is main focus ☐

Attention management: Comments here

Not present ☐
Attention management included ☐
Attention management is main focus ☐

Mindfulness / Meditation Comments here
Not present ☐
M/M is part of treatment ☐
M/M is main treatment focus ☐

Coping skills training Comments here
Not present ☐
CS is part of treatment ☐
CS is main treatment focus ☐

Cognitive restructuring: Comments here
Not present ☐
Cog restruct is part of treatment ☐
Cog restruct is treatment focus ☐

Spouse/Family involvement: Comments here
No spouse or family involvement ☐
Spouse or family member present for some part ☐
Spouse or family member involved as integral part of therapy ☐

Notes of on any other aspect of Treatment that may need to be discussed

Outcomes *** Use separate sheet for each measure***
Name of measure as used by author: _____
Domain of measurement

1. Coping/Cognitive appraisal ☐
2. Mood / affect ☐
3. Disability ☐
4. Pain behaviour ☐
5. Pain experience ☐
6. Social role functioning ☐
7. Health and social care use ☐
8. Physiology / fitness ☐

9. Other _____

Source of measurement:
Patient self-rating ☐
Spouse or family member ☐
Researcher / therapist, not blind to treatment ☐
Researcher, blind ☐
not stated ☐
Page number where data obtained: _____

Treatment/control name..... Treatment/control name.....

Mean pre-treatment:	_____	Mean pre-treatment:	_____
SD pre-treatment:	_____	SD pre-treatment:	_____
N pre-treatment:	_____	N pre-treatment:	_____
Mean post treatment:	_____	Mean post treatment:	_____
SD post treatment:	_____	SD post treatment:	_____
N post treatment:	_____	N post treatment:	_____
Mean f-up 1:	_____	Mean f-up 1:	_____
SD f-up 1:	_____	SD f-up 1:	_____
N f-up 1:	_____	N f-up 1:	_____
Mean f-up 2:	_____	Mean f-up 2:	_____
SD f-up 2:	_____	SD f-up 2:	_____
N f-up 2:	_____	N f-up 2:	_____

Treatment/control name..... Treatment/control name.....

Mean pre-treatment:	_____	Mean pre-treatment:	_____
SD pre-treatment:	_____	SD pre-treatment:	_____
N pre-treatment:	_____	N pre-treatment:	_____
Mean post treatment:	_____	Mean post treatment:	_____
SD post treatment:	_____	SD post treatment:	_____
N post treatment:	_____	N post treatment:	_____
Mean f-up 1:	_____	Mean f-up 1:	_____
SD f-up 1:	_____	SD f-up 1:	_____
N f-up 1:	_____	N f-up 1:	_____
Mean f-up 2:	_____	Mean f-up 2:	_____
SD f-up 2:	_____	SD f-up 2:	_____
N f-up 2:	_____	N f-up 2:	_____

Please give the times of the f-up data e.g. 6 months. _____

If the data is not in the form above is there a t or F test available if so what page is this on?

_____

Do the authors report either of the following?

t of pre-post change	Yes	☐	No	☐	on page no. _____
correlation of pre-post change	Yes	☐	No	☐	on page no. _____
Reliable change index	Yes	☐	No	☐	on page no. _____
Clinically significant change	Yes	☐	No	☐	on page no. _____

Notes of on any other aspect of the Outcome extraction that may need to be discussed

Appendix 5: Data Extraction Code Book

Value		Label		
arm_code	1	ACT as experiment		
	2	ACT as TAU		
	3	Behaviour therapy as experiment		
	4	Behaviour therapy as TAU		
	5	Biofeedback as experiment		
	6	Biofeedback as TAU		
	7	CBT as experiment (group) OUTPATIENT		
	8	CBT as TAU		
	9	Mindfulness as experiment		
	10	Mindfulness as TAU		
	11	Mixed as experiment		
	12	Mixed as TAU		
	13	Relaxation as experiment		
	14	Relaxation as TAU		
	15	education/bibliotherapy as experiment		
	16	education/bibliotherapy as control		
	17	arthritis education as experiment		
	18	arthritis education as TAU		
	19	specified standard care/TAU		
	20	Waiting list control		
	21	CBT as experiment (individual)		
	22	Exercise as experiment		
	23	Behavioural/Exercise as experiment		
	24	Cognitive therapy as experiment		
	25	cognitive & relaxation as experiment		
	26	CBT as experiment (GROUP) INPATIENT		
	27	Social support as experiment		
	28	Group CBT with family support (outpatient)		
	29	Education with family support		
	30	Spouse assisted group outpatient CBT		
	31	CBT plus biofeedback as experiment		
	32	Symptom monitoring and support as		

Value		Label		
		experiment		
	33	Individual physiotherapy (outpatient)		
	34	Group outpatient loving kindness meditation		
	35	Education and behavioural as experiment group outpatient		
	36	Behavioural plus physical plus CBT group outpatient		
	37	GP medical review plus education		
	38	Education/attention/support in group outpatient as control		
	39	Mixed individual and group CBT as intensive outpatient experiment		
	40	CBT as experiment (individual) plus psychoeducation and TAU		
	41	Individual TAU plus psychoeducation		
	42	Behaviour therapy as experiment (inpatient)		
	43	CBT as experiment (group inpatient) plus gender specific components		
	44	CBT group outpatient plus graded exposure		
		(note: C&CA = coping and cognitive appraisal, EF = emotional functioning, PF = physical functioning, Pain = pain experience, PB = pain behaviour)	OUTCOME DOMAIN	
measurement_code	1	coping strategies questionnaire (coping domain)	C&CA	0
	2	AIMS overall psychological dysfunction (mood domain)	EF	0
	3	AIMS anxiety (mood domain)	EF	0
	4	AIMS depression (mood domain)	EF	0
	5	HADS total (mood domain)	EF	0
	6	HADS anxiety (mood domain)	EF	0
	7	HADS depression (mood domain)	EF	0
	8	BAI (mood domain)	EF	0
	9	BDI (mood domain)	EF	0
	10	Brief symptom inventory (BSI) distress (mood domain)	EF	0

Value		Label		
	11	Centre for Epidemiologic Studies Depression Scale (CES-DS) (mood domain)	EF	0
	12	Clinical Outcomes in Routine Evaluation (CORE) total (mood domain)	EF	0
	13	CORE 34 W (mood domain)	EF	0
	14	CORE 34 P (mood domain)	EF	0
	15	CORE 34 F (mood domain)	EF	0
	16	CORE 34 R (mood domain)	EF	0
	17	Depression Adjective Checklist (DACL) (mood domain)	EF	0
	18	Depression Scale (DEPS) (mood domain)	EF	0
	19	Geriatric Depression Scale (GDS): (mood domain)	EF	0
	20	Hopkins Symptom Checklist Distress (HSCD): (mood domain)	EF	0
	21	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) Depression subscale (mood domain)	EF	0
	22	Minnesota Multiphasic Personality Inventory (MMPI) Depression subscale (mood domain)	EF	0
	23	State-Trait Anxiety Inventory State (STAI-S) Subscale (mood domain)	EF	0
	24	State-Trait Anxiety Inventory Trait (STAI-T) Subscale (mood domain)	EF	0
	25	(West Haven) Multidimensional Pain Inventory (WHYMPI) affective Distress (negative mood) subscale (mood domain)	EF	0
	26	Arthritis Impact Measurement Scale (AIMS) Activities of Daily Living (disability domain)	PF	0
	27	AIMS Physical Disability(disability domain)	PF	0
	28	AIMS mobility (disability domain)	PF	0
	29	AIMS dexterity (disability domain)	PF	0
	30	AIMS household activity (disability	PF	0

Value		Label		
		domain)		
	31	AIMS physical activity(disability domain)	PF	0
	32	Disability Rating Index (disability domain)	PF	0
EXCLUDE	33	Dusseldorf Disability Scale (disability domain) 6.4 DECIDED TO EXCLUDE AS NO DIRECTION POSSIBLE	PF	DK
	34	Fibromyalgia Impact Questionnaire total (disability domain)	PF	0
	35	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) mobility/Functional Disability (disability domain)	PF	0
	36	Oswestry Disability Index (disability domain)	PF	0
	37	Pain Disability Inventory (disability domain)	PF	0
	38	Roland and Morris Disability Questionnaire (disability domain)	PF	0
	39	Sickness Impact Profile (SIP; other) (disability domain)	PF	0
	40	Sickness Impact Profile (SIP; self) (disability domain)	PF	0
	41	SF-36 Physical Function/PHYSICAL HEALTH (disability domain)	PF	1
	42	(West Haven) Multidimensional Pain Inventory (WHYMPI) Activity subscale (disability domain)	PF	1
	43	(West Haven) Multidimensional Pain Inventory (WHYMPI) Pain Interference subscale (disability domain)	PF	0
	44	Pain Behaviour Direct observation (pain behaviour domain)	PB	0

Value		Label		
	45	Pain Behaviour Exercise tests (e.g. five or ten minute walk, sit to stand, stairs climbed in x minutes etc) (pain behaviour domain)	PF	1
	46	Pain behaviour Medication count (health and social care use)	PB	0
EXCLUDE	47	Pain behaviour Self-report schedule (pain behaviour domain) 6.4 DECIDED TO EXCLUDE AS NO DIRECTION POSSIBLE – MUST HAVE ONLY BEEN USED IN BOS ANYWAY AS NOT IN SPSS	PB	DK
	48	Arthritis Impact Measurement Scale (AIMS) pain subscale (pain experience domain)	Pain	0
	49	Arthritis Self-Efficacy Scale (pain experience domain)	C&CA	1
	50	Brief Pain Inventory (BPI) (pain experience domain)	Pain	0
	51	Characteristic Pain Intensity (pain experience domain)	pain	0
	52	Graded Chronic Pain Scale (pain experience domain)	Pain	0
	53	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) pain subscale (pain experience domain)	pain	0
	54	McGill Pain Questionnaire Pain Rating Index (pain experience domain)	Pain	0
	55	Memorial Pain Questionnaire (pain experience domain)	Pain	0
	56	(West Haven) Multi-dimensional Pain Inventory (MPI) Pain Severity/Intensity Rating (pain experience domain)	Pain	0
	57	Numerical Rating Scale (pain experience domain)	pain	0
	58	SF-36 Bodily Pain Scale (pain experience domain)	pain	1

Value		Label		
	59	Verbal rating pain scale (pain experience domain)	pain	0
	60	Visual Analogue Scale pain rating (pain experience domain)	pain	0
	61	Spielberger State Trait Anxiety Inventory (state subscale) (mood domain)	EF	0
	62	Sickness Impact Profile (SIP; other) (disability domain)	PF	0
	63	Sickness Impact Profile (SIP; self) (disability domain)	PF	0
	64	Visual Analogue Scale of Interference to social role functioning (social role fn domain)	PF	0
	65	(West Haven) Multi-dimensional Pain Inventory (MPI) Life Control Subscale coping/cognitive appraisal domain)	C&CA	1
	66	AIMS social activity subscale (social role fn domain)	PF	0
EXCLUDE	67	employment status (health and social care use domain)	TBC	EX
	68	Number of visits to GP (health and social care use domain)	PB	0
	69	Number of visits to other medical settings (health and social care use domain)	PB	0
	70	Medication Use (health and social care use domain)	PB	0
	71	Direct observation of motor pain behaviour (pain behaviour domain)	PB	0
	72	PSEQ (pain self-efficacy questionnaire; coping/cog appraisal domain)	C&CA	1
	73	Pain Catastrophising Scale (PCS; coping/cog appraisal domain)	C&CA	0
	74	chronic pain acceptance questionnaire (coping/cog appraisal domain)	C&CA	1
	75	DAPOS depression subscale (depression, anxiety and positive outlook scale; mood domain)	EF	0

Value		Label		
	76	DAPOS anxiety subscale (depression, anxiety and positive outlook scale; mood domain)	EF	0
	77	DAPOS positive outlook subscale (depression, anxiety and positive outlook scale; mood domain)	EF	1
	78	PASS - cognitive (Pain Anxiety Symptoms Scale, cognitive subscale, mood domain)	C&CA	0
	79	PASS - fear (Pain Anxiety Symptoms Scale, fear subscale, mood domain)	C&CA	0
	80	PASS - escape avoidance (Pain Anxiety Symptoms Scale, escape avoidance subscale, pain behaviour domain)	C&CA	0
	81	PASS - physiological (Pain Anxiety Symptoms Scale, physiological subscale, mood domain)	C&CA	0
	82	PASS total OR total short-form (Pain Anxiety Symptoms Scale, total, mood domain)	EF	0
	83	Brief Pain Inventory (BPI) interference subscale (disability domain)	PF	0
	84	non veterans Pain Outcome Questionnaire total (disability domain)	PF	0
	85	non veterans Pain Outcome Questionnaire ADL subscale (disability domain)	PF	0
	86	non veterans Pain Outcome Questionnaire mobility subscale (disability domain)	PF	0
	87	non veterans Pain Outcome Questionnaire vitality subscale (disability domain)	PF	0
EXCLUDE	88	non veterans Pain Outcome Questionnaire fear subscale (mood domain)	EF	0
EXCLUDE	89	non veterans Pain Outcome Questionnaire negative affect subscale (mood domain)	EF	0

Value		Label		
EXCLUDE	90	SF-8 health related QoL (total:Disability domain)	PF	1
EXCLUDE	91	WHO-QoL (total:disability domain)		dk
	92	Self-designed Physical Activity Profile (DIASBILITY domain)	PF	0
EXCLUDE	93	Self-designed Medication Review Questionnaire (health and social care use)		dk
EXCLUDE	94	British Pain Society Pain Rating Scale (pain experience domain)	Pain	0
	95	McGill Pain Questionnaire Short Form Pain Rating Index (pain experience domain)	Pain	0
	96	Tampa Scale for Kinesophobia total (coping/cognitive appraisal)	C&CA	0
EXCLUDE	97	Patient generated index of social role fn (social role fn domain)		dk
EXCLUDE	98	Satisfaction with support scale (social role fn domain)		dk
EXCLUDE	99	Euro-QoL (total: disability domain)		dk
EXCLUDE	100	Chronic pain coalition 5th vital sign scale (VAS; 1st 4 questions: pain experience domain)		dk
EXCLUDE	101	Chronic pain coalition 5th vital sign scale (VAS; last question: disability domain)		dk
EXCLUDE	102	Jenkins Sleep Scale (disability domain)	PF	0
	103	Fear Avoidance Beliefs Questionnaire (FABQ; coping/cognitive appraisal domain)	C&CA	0
EXCLUDE	104	Readiness to change questionnaire (coping/cog appraisal domain)		dk
EXCLUDE	105	Locus of control (coping/cog appraisal domain)		dk
EXCLUDE	106	Pain perception VAS (pain experience domain)		dk
EXCLUDE	107	Pain anger domain VAS (mood domain)		dk

Value		Label		
	108	Patient-Related Self-Statements (PRSS) catastrophising scale (coping/cognitive appraisal domain)	C&CA	0
	109	Patient-Related Self-Statements (PRSS) active coping scale (coping/cognitive appraisal domain)	C&CA	1
EXCLUDE	110	Pain Coping Inventory (coping/cognitive appraisal domain) 6.4 DECIDED TO EXCLUDE AS NO DIRECTION POSSIBLE – MUST HAVE ONLY BEEN USED IN BOS ANYWAY AS NOT IN SPSS	C&CA	dk
	111	Pain Coping Inventory, distraction by pleasant activities subscale (coping/cognitive appraisal domain)	C&CA	1
	112	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) self-care domain (disability domain)	PF	1
	113	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) anxiety subscale (mood domain)	EF	0
	114	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) potential emotional support subscale (social role functioning domain)	PF	1
	115	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) actual emotional support subscale (social role functioning domain)	PF	1
	116	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) mutual visits subscale (social role functioning domain)	PF	1
	117	Pain Behavior Checklist (pain behaviour domain)	PB	0
	118	Pain Behavior Checklist – modified (Spouse) (pain behaviour domain)	PB	0
N/B IN FINAL MEASURE COUNT JUST COUNT THESE WITH NORMAL SIP FOLLOWING	119	Sickness Impact Profile modified (SIP; self) (disability domain)	PF	0

Value		Label		
15.1.2010 DISCUSSION WITH SM				
N/B IN FINAL MEASURE COUNT JUST COUNT THESE WITH NORMAL SIP FOLLOWING 15.1.2010 DISCUSSION WITH SM	120	Sickness Impact Profile modified (SIP; other) (disability domain)	PF	0
	121	Cognitive Errors Questionnaire (coping/cognitive appraisal domain)	C&CA	1
	122	Pain Disability Index total score (disability domain)	PF	0
	123	Pain Beliefs Questionnaire (coping/cognitive appraisal domain)	C&CA	0
	124	Pain diary pain experience (pain experience domain)	Pain	0
	125	Tubingen Pain Behavior Scale (PBS; pain behaviour domain)	PB	0
EXCLUDE	126	Electromyographic (EMG) baseline (microvolts) (physiology/fitness domain)		EX
EXCLUDE	127	Electromyographic EMG change from baseline (microvolts) (physiology/fitness domain)		EX
	128	Likert scale pain interference (disability domain)	PF	0
	129	Likert scale pain distress (pain experience)	pain	0
	130	Likert pain rating (pain experience domain)	pain	0
	131	Visual analogue scale pain experience SPOUSE (pain experience domain)	Pain	0
	132	Pain Cognitions Questionnaire – HOPELESSNESS SUBSCALE (PCQ; coping/cognitive appraisal domain)	C&CA	0
	133	State-Trait Anxiety Inventory TOTAL (STAI) (mood domain)	EF	0
	134	Pain Cognitions Questionnaire – ACTIVE COPING (PCQ; coping domain)	C&CA	1

Value		Label		
	135	Non-specified self-efficacy scale (coping/cognitive appraisal domain)	C&CA	1
	136	Lehmann low back physical scale (physical subscale (disability domain)	PF	1
	137	Lehmann low back pain rating scale (patient reported disability; disability domain)	PF	1
	138	Lehmann low back pain rating scale (physician reported disability; disability domain)	PF	1
	139	Lehmann low back pain rating scale (total; disability domain)	PF	1
	140	Check list for inter-personal pain behaviour: distorted mobility scale (CHIP distorted mobility; pain behaviour domain)	PF	0
	141	Check list for inter-personal pain behaviour: nervousness scale (CHIP nervousness; mood domain)	EF	0
	142	Check list for inter-personal pain behaviour: depression scale (CHIP depression; mood domain)	EF	0
	143	Pain Cognitions List: impact (PCL impact; coping/cognitive appraisal domain)	C&CA	0
	144	Pain Cognitions List: catastrophising (PCL catastrophising; coping/cognitive appraisal domain)	C&CA	0
	145	Behavioural Approach Test walking distance (BAT walking distance; pain behaviour domain)	PF	1
	146	Behavioural Approach Test walking time (BAT walking time; pain behaviour domain)	PF	1
	147	Pain Cognitions List: outcome efficacy (PCL outcome efficacy; coping/cognitive appraisal domain)	C&CA	0
	148	Likert coping (coping/cognitive appraisal domain)	C&CA	1
	149	Pain Management Inventory: active coping (PMI active; coping/cognitive	C&CA	0

Value		Label		
		appraisal domain)		
	150	Pain Management Inventory: passive coping (PMI passive; coping/cognitive appraisal domain)	C&CA	0
	151	Combined measures of pain control (coping/cognitive appraisal)	C&CA	1
	152	Coping strategies questionnaire; relaxation (coping/cognitive appraisal domain)	C&CA	1
	153	Combined measures of catastrophising (coping/cognitive appraisal)	C&CA	0
	154	Combined measures of pain intensity (pain experience domain)	Pain	0
	155	Combined measures of pain behaviour(pain behaviour domain)	Pain	0
	156	Combined measures of activity (walking distances etc (disability domain)	PF	1
	157	Maudsley Obsessive Compulsive Inventory (unclear domain)	EF	0
	158	Fear Survey Schedule	EF	0
EXCLUDE	159	Dutch Hyperventilation Questionnaire (DHQ; PAIN BEHAVIOUR DOMAIN)		ex
	160	Coping strategies coping scale (coping/cognitive appraisal domain)	C&CA	1
	161	Coping strategies pain control scale (coping/cognitive appraisal domain)	C&CA	1
	162	Health assessment questionnaire (disability domain)	PF	0
	163	Arthritis self-efficacy scale function (coping/cognitive appraisal)	C&CA	1
	164	Arthritis self-efficacy scale – arthritis (coping/cognitive appraisal)	C&CA	1
	165	Zung depression scale (mood)	EF	0
	166	Perceived stress scale (coping/cognitive appraisal)	C&CA	0
	167	Perceived stress scale – coping scale (coping/cognitive appraisal)	C&CA	1

Value		Label		
	168	UCLA loneliness scale (coping/cognitive appraisal domain)	C&CA	0
	169	SF-36 Physical Role Function (disability domain)	PF	1
	170	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) social network (social role functioning domain)	PF	1
	171	Checklist individual strength fatigue scale (physiology/fitness)	PF	0
	172	Pain Coping Inventory ACTIVE COPING (coping/cognitive appraisal domain)	C&CA	1
	173	Pain Coping Inventory PASSIVE COPING (coping/cognitive appraisal domain)	C&CA	0
	174	Illness cognition questionnaire helplessness(coping/cognitive appraisal do	C&CA	0
	175	Illness cognition questionnaire acceptance (coping/cognitive appraisal domain)	C&CA	1
	176	Utrechtse Coping Lijst stress active coping (coping/cognitive appraisal domain)	C&CA	1
	177	Utrechtse Coping Lijst stress passive coping (coping/cognitive appraisal domain)	C&CA	0
	178	Impact of Rheumatic diseases on General health and Lifestyle (IRGL) mobility and self-care composite score (disability domain)	PF	0
	179	Perceived stress scale 4 item version (coping/cognitive appraisal)	C&CA	0
	180	SF-36 fatigue severity (disability domain)	PF	1
	181	SF-36 global self-assessment (disability domain)	PF	1
	182	likert subjective working capacity (coping/cognitive appraisal)	C&CA	1
	183	likert benefits of working after 2 years(coping/cognitive appraisal)	C&CA	1
	184	likert general wellbeing (disability domain)	PF	0
	185	Fibromyalgia Impact Questionnaire Physic	PF	0

Value		Label		
		function (disability)		
	186	Fibromyalgia Impact Questionnaire feel good (mood)	EF	0
	187	Fibromyalgia Impact Questionnaire VAS pain (pain experience)	Pain	0
	188	Fibromyalgia Impact Questionnaire VAS fatigue (disability)	PF	0
	189	Fibromyalgia Impact Questionnaire VAS sleep (disability)	PF	0
	190	Fibromyalgia Impact Questionnaire VAS stiffness (disability)	PF	0
	191	Fibromyalgia Impact Questionnaire VAS anxiety (mood)	EF	0
	192	Fibromyalgia Impact Questionnaire VAS depression (mood)	EF	0
	193	McGill pain rating index intensity subscale (pain experience domain)	pain	0
	194	SF-36 general health (disability domain)	PF	1
	195	SF-36 vitality (disability)	PF	1
	196	SF-36 social functioning (social role functioning)	PF	1
	197	SF-36 emotional role (disability)	PF	1
	198	SF-36 mental health (mood/affect)	EF	1
	199	Chronic Pain Self-Efficacy Scale pain management (coping/cognitive appraisal)	C&CA	1
	200	SF-36 health change (coping/cognitive appraisal)	C&CA	1
	201	Chronic Pain Self-Efficacy Scale physical function (coping/cognitive appraisal)	C&CA	1
	202	Chronic Pain Self-Efficacy Scale symptoms (coping/cognitive appraisal)	C&CA	1
	203	Chronic Pain Coping Inventory asking for assistance (coping/cognitive appraisal)	C&CA	0

Value		Label		
	204	Chronic Pain Coping Inventory guarding (coping/cognitive appraisal)	C&CA	0
	205	Chronic Pain Coping Inventory resting (coping/cognitive appraisal)	C&CA	0
	206	Chronic Pain Coping Inventory relaxation (coping/cognitive appraisal)	C&CA	1
	207	Chronic Pain Coping Inventory task persistence (coping/cognitive appraisal)	C&CA	1
	208	Chronic Pain Coping Inventory exercise (coping/cognitive appraisal)	C&CA	1
	209	Chronic Pain Coping Inventory social support (coping/cognitive appraisal)	C&CA	1
	210	Chronic Pain Coping Inventory self-statements (coping/cognitive appraisal)	C&CA	1
	211	McGill pain rating index sensory subscale (pain experience domain)	Pain	0
	212	McGill pain rating index affective subscale (pain experience domain)	Pain	0
	213	McGill pain rating index evaluative subscale (pain experience domain)	Pain	0
	214	Brief pain inventory usual pain (pain experience domain)	Pain	0
	215	Brief pain inventory worst pain (pain experience domain)	Pain	0
	216	State trait anger expression STAXI-II anger expression in(Mood domain)	EF	0
	217	State trait anger expression STAXI-II anger expression out (Mood domain)	EF	0
	218	State trait anger expression STAXI-II control in (Mood domain)	EF	0
	219	State trait anger expression STAXI-II control out (Mood domain)	EF	0
	220	AIMS2 upper limb (disability)	PF	0
	221	AIMS2 lower limb (disability)	PF	0
	222	Arthritis Self-Efficacy Scale OTHER symptoms (self-efficacy domain)	C&CA	1
	223	Rheumatoid Attitudes Index Helplessness (coping/cognitive appraisal)	C&CA	0

Value		Label		
	224	Rheumatoid Attitudes Index Internality(coping/cognitive appraisal)	C&CA	0
	225	AIMS2 health status (ie AIMS total) (disability domain)	PF	0
	226	AIMS2 satisfaction with health (domain TBC)	PF	0
	227	Joint Protection Behaviour Assessment (JPBA) (pain behaviour)	PB	1
	228	Hand pain visual analogue scale (pain experience)	Pain	0
	229	Graded Chronic Pain Scale activity interference (disability domain)	PF	0
	230	Graded Chronic Pain Scale pain intensity (pain experience domain)	Pain	0
	231	Survey of Pain Beliefs (SOPA) disability (disability domain)	C&CA	0
	232	Survey of Pain Beliefs (SOPA) harm (Coping/cognitive appraisal)	C&CA	0
	233	Survey of Pain Beliefs (SOPA) control (coping/cognitive appraisal)	C&CA	1
	234	TMD Self-Efficacy Scale (COPING/COGNITIVE APPRAISAL)	C&CA	1
	235	Coping strategies questionnaire catastrophising domain (coping/cognitive appraisal)	C&CA	0
	236	Pain Catastrophising Scale rumination (PCS; coping/cog appraisal domain)	C&CA	0
	237	Days without pain per week (pain experience)	Pain	1
	238	Likert control over pain (coping/cognitive appraisal)	C&CA	1
	239	Days with medication per week (health and social care use)	PB	0
	240	Heidelberg Coping Scale Cognitive Strategies (coping/cognitive appraisal)	C&CA	1
	241	Heidelberg Coping Scale avoidance behaviour (Coping/cognitive appraisal)	C&CA	0
	242	Heidelberg Coping Scale pleasant activities (Coping/cognitive appraisal)	C&CA	1

Value		Label		
	243	Heidelberg Coping Scale social support (Coping/cognitive appraisal)	C&CA	1
	244	Heidelberg Coping Scale philosophical beliefs (coping/cognitive appraisal)	C&CA	0
	245	Dusseldorf Disability Scale disability to social relations (social role functioning)	PF	0
	246	Dusseldorf Disability Scale disability to social roles (social role functioning)	PF	0
	247	Dusseldorf Disability Scale disability to physical functions (disability)	PF	0
	248	Dusseldorf Disability Scale disability to mental performance (disability)	PF	0
	249	Dusseldorf Disability Scale disability to physical performance (disability)	PF	0
	250	Psychological General Well Being scale (disability domain)	PF	1
	251	Shortened version (16 item) Roland and Morris Disability Questionnaire (disability domain)	PF	0
	252	Physician related concerns re pain meds intake (health and social care use domain)	PB	0
	253	Brief Norwegian version of Hopkins Symptom Checklist : (mood domain)	EF	0
	254	Ursin's health inventory subjective health total (disability domain)	PF	0
	255	Visual Analogue Scale of Interference (disability domain)	PF	0
	256	(West Haven) Multidimensional Pain Inventory (WHYMPI) household chores (disability domain)	PF	1
	257	(West Haven) Multidimensional Pain Inventory (WHYMPI) outdoor work (disability domain)	PF	1
	258	(West Haven) Multidimensional Pain Inventory (WHYMPI) activities away from home (disability domain)	PF	1
	259	(West Haven) Multidimensional Pain Inventory (WHYMPI) social activities(disability domain)	PF	1

Value		Label		
	260	Coping strategies questionnaire; diverting attention (coping/cognitive appraisal domain)	C&CA	1
	261	Coping strategies questionnaire; reinterpret pain sensations (coping/cognitive appraisal domain)	C&CA	1
	262	Coping strategies questionnaire; coping self-statement (coping/cognitive appraisal domain)	C&CA	1
	263	Coping strategies questionnaire; ignore pain sensations (coping/cognitive appraisal domain)	C&CA	1
	264	Coping strategies questionnaire; praying and hoping (coping/cognitive appraisal domain)	C&CA	0
	265	Coping strategies questionnaire; increase activity level (coping/cognitive appraisal)	C&CA	1
	266	Coping strategies questionnaire; increase pain behaviours (pain behaviour domain)	C&CA	0
	267	(West Haven) Multidimensional Pain Inventory (WHYMPI) support (social role functioning domain)	PF	1
	268	(West Haven) Multidimensional Pain Inventory (WHYMPI) punishing responses (negative responses) (social role functioning)	PF	0
	269	(West Haven) Multidimensional Pain Inventory (WHYMPI) solicitous responses (social role functioning)	PF	0
	270	(West Haven) Multidimensional Pain Inventory (WHYMPI) distracting responses (social role functioning)	PF	0
EXCLUDE	271	Pain Discomfort Scale (pain domain)		0
EXCLUDE	272	11-Point Box Scale (likert) (pain experience domain)		0
	273	Quality of Well Being Scale (QWB) (disability domain)	PF	1
	274	Myalgia score (pain experience domain)	pain	0
	275	Quality of social support scale (social role	PF	1

Value		Label		
		functioning)		
	276	Combined measures of depressed and negative cognitions (mood domain)	EF	0
EXCLUDE	277	Combined measures of using acute pain strategies (coping/cognitive appraisal) 6.4 DECIDED TO EXCLUDE AS NO DIRECTION POSSIBLE – MUST HAVE ONLY BEEN USED IN BOS ANYWAY AS NOT IN SPSS	C&CA	
	278	(West Haven) Multidimensional Pain Inventory (WHYMPI) self-efficacy scale – item added to German version (coping/cognitive appraisal)	C&CA	1
	279	No. Of hours sleep (disability domain)	PF	1
	280	Visual analogue scale anxiety (mood domain)	EF	0
	281	Global self rating index of health perception (GSI; disability)	PF	0
	282	Arm endurance test (physical functioning domain)	PF	1
	283	Sleep quality (physical functioning)	PF	1
	284	Survey of Pain Beliefs (SOPA) medical cure (coping/cognitive appraisal)	C&CA	0
	285	Upper limb activity (physical functioning)	PF	1
	286	Lower limb activity (Physical functioning domain)	PF	1
	287	Vertebral column activity (Physical functioning domain	PF	1
	288	Mandibular function impairment questionnaire masticatory(MFIQ; physical function)	PF	0
	289	Mandibular function impairment questionnaire non-masticatory(MFIQ; physical function)	PF	0
	290	Sleep quality where decrement = improvement	PF	0

Value		Label		
	291	Brief Pain Inventory (BPI) intensity subscale (disability domain)	PF	0
	292	Modified 8-item Arthritis Self-Efficacy Scale total (coping/cognitive appraisal)	C&CA	1
	293	Chronic Pain Coping Inventory pacing(coping/cognitive appraisal)	C&CA	1
	294	Quebec back pain disability scale (any language version) disability	PF	0
	295	Patient specific complaints (visual analogue scale) (disability)	PF	0
	296	Short version Photograph Series of Daily Activities (PHODA-SeV) (coping/cognitive appraisal)	PF	1
EXCLUDE	297	RT3 Tri-Axial Research Tracker (activity levels) 6.4 DECIDED TO EXCLUDE	PF	1
	298	chronic pain acceptance questionnaire – activity engagement subscale (coping/cognitive appraisal)	C&CA	1
	299	Positive and negative affect schedule (PANS) positive affect subscale	EF	1
	300	Positive and negative affect schedule (PANS) negative affect subscale	EF	0
	301	Pain self-efficacy scale (PSEQ; coping cognitive appraisal)	C&CA	1
	302	Depression and Anxiety Stress Scales (DASS) depression subscale	EF	0
	303	Depression and Anxiety Stress Scales (DASS) anxiety subscale	EF	0
	304	Arthritis self-efficacy scale overall (coping/cognitive appraisal)	C&CA	1
	305	Dyadic adjustment scale (role functioning)	PF	1
	306	Minnesota Multiphasic Personality Inventory (MMPI) Somatization subscale (domain)	EF	0
	307	PARS (Spouse related community adjustment	PF	1
NB. 278 separate outcome measures were used in the final pre-post analysis once those				

Value		Label		
whose direction could not be ascertained had been removed and duplicates merged.				
Measurement_direction	0	decrement is improvement		
	1	increment is improvement		
measurement_domain	1	coping/cognitive appraisal		
	2	mood		
	3	disability		
	4	pain behaviour (includes beh. activity, may do post hoc analysis)		
	5	pain experience		
	6	social role functioning		
	7	health and social care use		
	8	(physiology/fitness - unlikely to use)		
	9	unclear		
IMMPACT/coping cog appraisal domain	1	Pain experience		
	2	Physical functioning		
	3	Emotional functioning		
	4	Coping and cognitive appraisal		
	5	TBC		
	6	Pain behaviour		
measurement_source	1	patient		
	2	spouse/family		
	3	researcher blind		
	4	researcher not blind		
	5	not stated		
reactivity_measure	0	low (eg patient rated)		
	1	high (eg non patient rated)		
	2	unclear		
specficity_measure	0	low (eg not related to target problem)		
	1	high (eg related to target problem)		
	2	unclear		
effect_sizes_recorded	0	no		
	1	yes		
	2	unclear		
ITT_analyses_conducted	0	no		

Value		Label		
	1	yes		
	2	unclear		
RCI_calculated	0	no		
	1	yes		
	2	unclear		
t_of_pre_post_change				
correlation_pre_post_change				
outcome_data_page_numbers				
mean_pre_treatment				
SD_pre_treatment				
n_pre_treatment				
mean_post_treatment				
SD_post_treatment				
n_post_treatment				
time_to_follow_up_1	1	1 week after post treatment		
	2	2 weeks after post treatment		
	3	3 weeks after post treatment		
	4	4 weeks after post treatment		
	5	5 weeks after post treatment		
	6	6 weeks after post treatment		
	12	12 weeks after post-treatment		
	14	14weeks after post treatment		
	16	4 months after post treatment		
	26	6 months after post treatment		
	36	9 months post-treatment		
	52	Year after post treatment		
	60	15 months post treatment		
	104	2 year after post treatment		
mean_follow_up_1				
SD_follow_up_1				
n_follow_up_1				
time_to_follow_up_2	1	1 week after post treatment		
	2	2 weeks after post treatment		
	3	3 weeks after post treatment		

Value		Label		
	4	4 weeks after post treatment		
	5	5 weeks after post treatment		
	6	6 weeks after post treatment		
	12	12 weeks after post treatment		
	14	14weeks after post treatment		
	16	4 months after post treatment		
	26	6 months after post treatment		
	36	9 months post treatment		
	52	Year after post treatment		
	60	15 months post treatment		
	78	18 months post treatment		
	104	2 year after post treatment		
mean_follow_up_2				
SD_follow_up_2				
n_follow_up_2				
(same as above for follow up 3 too)				

Appendix 6: Survey Round One, Outcome Domains (study 3)

In order that I am able to describe the composition of this expert panel, it would be helpful if you could complete the following 'about you' questions. These will not be connected to your subsequent survey responses in any way.

1. Please record your current profession and job title (if this is different) in the below space

2. How long ago did you achieve your professional qualification?

3. For how many years have you worked within this PMP?

4. Are you male or female?

Thank you for providing this information.

Survey 1: Assessing the importance of particular outcome domains

This survey asks you to consider a number of potential outcome domains that PMPs could employ. This list is based upon a number of suggestions within the research literature in addition to the collective responses within the previous PMP survey. At this stage I would like you to consider these domains and rate the importance of each one. I would also like you to consider and comment upon the definition of each domain which has been provided and, if applicable, suggest specific measures within each domain that you rate as important. I appreciate that not all measures will be of relevance to your professional role, but would like you to consider the multi-disciplinary nature of PMPs within your ratings and recommendations.

5. Potential Outcome Domain

a. Coping or cognitive appraisal

b. Mood

c. Disability

d. Pain behaviour or activity

e. Pain experience

f. Social role functioning

g. Physiology or fitness

Indicate whether you think it is:

very important

somewhat important

unimportant, or

definitely not important

to measure these domains within your PMP.

Please select the 'More Info' box (to the right) to view domain definitions, and use the space below to comment upon these.

If you have indicated that this domain is important, please use the below space to recommend specific measures.

Appendix 7: Survey Round Two, Outcome Domains (study 3)

In order that I am able to describe the composition of this expert panel, it would be helpful if you could complete the following 'about you' questions.

These will not be connected to your subsequent survey responses in any way.

1. Please record your current profession and job title (if this is different) in the below space

2. How long ago did you achieve your professional qualification?

3. For how many years have you worked within this PMP?

4. Are you male or female?

Thank you for providing this information.

The second survey

This survey asks you to consider the collective responses to the previous survey and suggest improvements to domain definitions. It also asks you to consider the collective ratings of importance assigned to these domains, and decide whether you agree that each domain should be included. Finally, it asks you to consider the outcome measures suggested in the previous survey, and assign each one to a domain. I appreciate that not all measures will be of relevance to your professional role, but would like you to consider the multi-disciplinary nature of PMPs within your responses.

Domain definitions and importance

5. Coping or cognitive appraisal: e.g. measures of cognitive strategies and appraisals used to manage pain.

One expert suggested that repeating the words 'cognitive' and 'appraisal' made this definition unclear. Others implied the importance of measures within this domain, including catastrophising, changes to thinking patterns as indicated by re-appraisal of the situation, self-efficacy and fear of movement.

Please use the below space to record your views upon these comments and how this domain could be defined.

In the previous survey, 91.7% of experts (n=11) rated this domain as very important, and 8.3% (n=1) as somewhat important.

With this in mind and your reflections upon the definition above, do you think that PMPs should definitely measure outcomes within this domain?

Yes, No

6. Mood: e.g. depressive state, anxious state.

Experts suggested that the examples used were useful but restrictive and that experiences such as anger and frustration should also be incorporated, with one suggestion that response to mood may be more significant than mood itself.

Please use the below space to record your views upon these comments and how this domain could be defined.

In the previous survey, two-thirds of experts (n=8) rated this domain as very important, with the remaining third (n=4) rating it as somewhat important.

With this in mind and your reflections upon the definition above, do you think that PMPs should definitely measure outcomes within this domain?

Yes, No

7. Disability: e.g. activities of daily living, impact on health and lifestyle.

Only one expert commented on the definition itself. Others suggested that it was fundamental to assess disability, with a contrasting view that PMP patients may have inaccurate views of the extent of their disability and therefore measuring it may not be meaningful.

Please use the below space to record your views upon these comments and how this domain could be defined.

In the previous survey, two-thirds (n=8) rated disability as very important to measure, with 25% (n=3) viewing it as somewhat important and 8.3% (n=1) as unimportant.

With this in mind and your reflections upon the definition as above, do you think that PMPs should definitely measure outcomes within this domain?

Yes, No

8. Pain behaviour or activity: e.g. behavioural acts associated with pain, including distance able to walk etc.

Several experts suggested that the amalgamation of pain behaviour and activity was logical, with a contrasting view that they should be classed separately. The view that the definition was too vague and should be more specific, perhaps including fear-avoidance related behaviours, help-seeking and healthcare use was also expressed, with one additional view that the level of importance assigned to the domain may be dependent on the domain definitions themselves.

Please use the below space to record your views upon these comments and how this domain could be defined.

In the previous survey, two-thirds of experts (n=8) rated this domain as 'very important, with 25% (n=3) and 8.3% (n=1) rating it as somewhat important and unimportant respectively.

With this in mind and your reflections upon the definition above, do you think that PMPs should definitely measure outcomes within this domain?

Yes, No

9. Pain experience: e.g. ratings of pain intensity, sensation and unpleasantness.

This was the domain which led to the least consensus. Experts commented that pain experience is hard to define due to its subjectivity, and therefore that it may be a less important domain to measure within PMPs, particularly as treatment progresses.

Please use the below space to record your views upon these comments and how this domain could be defined.

In the previous survey, 25% (n=3) rated this domain as very important, 50% (n=6) as somewhat important, 16.7% (n=2) and 8.3% (n=1) as unimportant and definitely not important respectively.

With this in mind and your reflections upon the definition above, do you think that PMPs should definitely measure outcomes within this domain?

Yes, No

10. Social Role Functioning: e.g. assessments of the ability of the person to function in various social roles, including familial, leisure and employment related roles.

There was a general view that it would be more appropriate to re-name this domain 'role' functioning given that the definition incorporated home and family as well as work and leisure roles. There was also an expectation that outcomes within this domain may overlap with those in the pain behaviour and coping domains, and be of variable levels of importance depending on individual patients.

Please use the below space to record your views upon these comments and how this domain could be defined.

In the previous survey, two-thirds (n=8) of experts viewed this as a very important domain to measure, with 25% (n=3) and 8.3% (n=1) viewing it is somewhat important and unimportant respectively.

With this in mind and your reflections upon the definition above, do you think that PMPs should definitely measure outcomes within this domain?

Yes, No

11. Physiology/fitness: e.g. assessments of biological functioning and physical fitness.

Several participants commented that timed walks may be applicable within this domain, and that it was a difficult domain to measure.

Please use the below space to record your views upon these comments and how this domain could be defined.

Participants in the previous survey varied greatly in terms of the importance they assigned to measures within this domain, with 25% (n=3) rating it as very important, 58.3% (n=7) as somewhat important and 16.7% (n=2) as unimportant.

With this in mind and your reflections upon the definition above, do you think that PMPs should definitely measure outcomes within this domain?

Yes, No

12. In the first survey you suggested a number of outcome measures that would be useful within PMPs (listed below) but there was a lack of consensus regarding the domains within which particular measures were assigned.

You are now invited to assign each measure that you are familiar with to the domain which you feel is most appropriate.

Some measures comprise scales which are relevant to more than one domain, i.e. are multi-dimensional. If you perceive a measure to be multi-dimensional, please select this and state the most appropriate domain for the total measure score, and suggest domains for the particular scales within the measure.

Please state the most appropriate domain for the total measure score and the scales within this measure.

N/A - I am not familiar with the measure

Coping or cognitive appraisal

Mood

Disability

Pain behaviour or activity

Pain experience

Social role functioning

Physiology or fitness

Multi-dimensional

a. Arm Rotation

b. Beck Anxiety Inventory

c. Beck Depression Inventory

d. Brief Pain Inventory

e. Canadian Occupational Performance Measure activity scale

f. Canadian Occupational Performance Measure total

g. Chronic Pain Acceptance Questionnaire

h. Coping Strategies Questionnaire

i. Fear Avoidance Beliefs Questionnaire

j. General Health Questionnaire-12

13.

Please state the most appropriate domain for the total measure score and the scales within this measure.

N/A - I am not familiar with the measure

Coping or cognitive appraisal

Mood

Disability

Pain behaviour or activity

Pain experience

Social role functioning

Physiology or fitness

Multi-dimensional

a. Hospital Anxiety and Depression Scale

b. McGill short-form

c. Medication Count

d. Nottingham Health Profile

e. Observation (non-specified)

f. Oswestry Disability Index

g. Pain Catastrophising Scale

h. Pain Disability Questionnaire

i. Pain Self-Efficacy Questionnaire

j. Roland Morris Disability Questionnaire

14. Please state the most appropriate domain for the total measure score and the scales within this measure.

N/A - I am not familiar with the measure

Coping or cognitive appraisal

Mood

Disability

Pain behaviour or activity

Pain experience

Social role functioning

Physiology or fitness

Multi-dimensional

a. Short-Form 8 total

b. Short-Form 36 total

c. Short- Form-36 Bodily Pain

d. Sickness Impact Profile

e. Sit and reach

f. Sit to stand

g. Tampa Scale for Kinesophobia

h. Visual Analogue Scale distress

i. Visual Analogue Scale interference

j. Walking

Appendix 8: Survey Round Two Summary (study 3)

Dear colleague,

In recent months you have been a member of an expert panel of thirteen whose initial task was to generate a list of outcome measures of use within Pain Management Programmes (PMPs) by completing electronic surveys; conducted in line with the Delphi method. The data from two surveys has now been analysed, and I am writing to you with a summary of the results and to confirm that this part of my research is now complete.

The results of the first survey demonstrated a broad consensus regarding the utility of specific outcome domains, but more variability with regard to the domain definitions and assignment of measures to domains; congruent within a recently published review of measures for people with chronic pain (reference supplied). The purpose of the second survey was therefore amended, and it sought more clarification regarding domain definitions and the assignment of measures.

Of the 77% of experts who completed the second survey (n=10) there was a general consensus (defined as 70% agreement) that measures within the domains of coping and cognitive appraisal, mood, disability and social role functioning should definitely be measured within PMPs. There was a lack of consensus, however, as to whether measures within the domains of pain behaviour, pain experience and physiology/fitness should be incorporated, and upon the most useful manner in which each domain could be defined. Overall, experts agreed that the value of a measure should be determined by the context and the specific aims of a PMP, and that overlap between domains should not preclude the use of any measure.

As a consequence of these results I have updated a review of the literature regarding outcome domains within chronic pain, and the results of both will ultimately inform the manner in which data from Randomised Controlled Trials (RCTs) is aggregated within the meta-analysis that I will conduct in the coming months.

I would like to take this opportunity to thank you for your participation; it has been much appreciated. If you have any further queries regarding this research, or would like a copy of my thesis when it is complete then please let me know and I will be happy to forward it to you.

Yours faithfully,

Grania Fenton (contact details supplied)