Article Type: Original Article

Clinical course and prognostic factors across different musculoskeletal pain sites: A secondary analysis of individual patient data from randomised clinical trials

Running head: Prognostic similarities in musculoskeletal pain

D.J. Green[1], M. Lewis[1], G. Mansell[1], M. Artus[1], K.S. Dziedzic[1], E.M. Hay[1], N.E. Foster[1], D.A. van der Windt[1]

[1] Arthritis Research UK Primary Care Centre (Research Institute for Primary Care & Health Sciences) & Keele Clinical Trials Unit (David Weatherall Building), Keele University, Keele, Staffordshire, ST5 5BG

Corresponding author: Danielle van der Windt,
Arthritis Research UK Primary Care Centre, Institute for Primary Care and Health Sciences, Keele University, Staffordshire, ST5 5BG
Email: d.van.der.windt@keele.ac.uk; Tel: +44 (0)1782 734830; Fax: +44 (0)1782 734719

Original article

The project was supported by a grant (project no. 84) from the National School for Primary Care (NIHR) School for Primary Care Research. DJG was funded by a NIHR School for Primary Care Research Doctoral Training Studentship; KD is part funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Research and Care West Midlands and by a Knowledge Mobilisation Research Fellowship (KMRF-2014-03-002) from the NIHR; NF, an NIHR Senior Investigator, is supported through an NIHR Research Professorship (NIHR-RP-011-015); DvdW is a member of PROGRESS Medical Research Council Prognosis Research Strategy (PROGRESS) Partnership (G0902393/99558); GM is supported by NIHR School for Primary Care Seedcorn Funding; EH is a Senior NIHR Investigator. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

The authors have no conflicts of interest to declare.

Significance : Individual patient data analysis of trials across different regional musculoskeletal pain sites was used to evaluate course and prognostic factors associated with pain and disability. Overall, similarity of outcome predictors across these different pain sites provides supports targeting of treatment based on prognostic factors rather than pain site alone.

ABSTRACT

Background: Previous research has identified similar prognostic factors in patients with musculoskeletal (MSK) conditions regardless of pain presentation, generating opportunities for management based on prognosis rather than specific pain presentation.

Methods: Data from seven RCTs (2,483 participants) evaluating a range of primary care interventions for different MSK pain conditions were used to investigate the course of symptoms and explore similarities and differences in predictors of outcome. The value of pain site for predicting changes in pain and function was investigated and compared with that of age, gender, social class, pain duration, widespread pain, and level of anxiety/depression.

Results: Over the initial three months of follow-up, changes in mean pain intensity reflected an improvement, with little change occurring after this period. Participants with knee pain due to osteoarthritis (OA) showed poorer long-term outcome (mean difference in pain reduction at 12 months -1.85, 95% CI -2.12 to -1.57, compared to low back pain). Increasing age, manual work, longer pain duration, widespread pain, and increasing anxiety/depression scores were significantly associated with poorer outcome regardless of pain site. Testing of interactions showed some variation between pain sites, particularly for knee OA, where age, manual work and pain duration were most strongly associated with outcome.

Conclusions: Despite some differences in prognostic factors for trial participants with knee OA who were older and had more chronic conditions, similarity of outcome predictors across regional MSK pain sites provides evidence to support targeting of treatment based on prognostic factors rather than site of pain.

INTRODUCTION

The majority of studies on prognosis and management of patients with musculoskeletal (MSK) pain have focussed on specific regional pain presentations, such as low back, shoulder, or knee pain. Many of these studies show similar findings in terms of clinical characteristics (van der Windt et al 2008), symptom trajectories (Henschke et al 2012), and prognostic factors (Artus et al 2017; van der Windt 2010; Henschke et al 2012). Systematic reviews of prognostic factors have consistently

identified pain duration and functional limitations as predictive of poor outcome in upper limb pain (Bruls et al 2015; Kooijman et al 2015), low back pain (LBP) (Chou & Shekelle 2010), and knee pain attributable to osteoarthritis (OA) (Bastick et al 2015). Observational research has identified socioeconomic variables, baseline pain characteristics, and psychological factors that consistently predict outcome regardless of pain site (e.g. Valentin et al 2016; Artus et al 2017; de Vos Andersen et al 2017). Similarly, prognostic scores for estimating risk of persistent disabling pain, developed originally in back pain patients (Von Korff & Miglioretti 2005; Hill et al 2016), have been shown to accurately predict outcome across a range of MSK pain sites (Thomas et al 2008; Mallen et al 2013). Finally, a brief set of generic prognostic factors (duration of pain episode, pain interference with daily activities, presence of multiple-site pain) was found to improve on clinicians' estimates of prognosis in older patients with a range of MSK presentations in primary care (Mallen et al 2013).

Localised, single-site pain is rare with 40-75% of individuals reporting pain at multiple sites (Carnes et al 2007; Kalameri et al 2008; van der Windt et al 2008; Hartvigsen et al 2013), which may partly explain similarities across pain presentations. A population-based survey (*n*=3,179) showed 53% of those with pain reported pain in more than one site (Kalameri et al 2008), with the number of pain sites strongly associated with reduced physical functioning, symptoms of anxiety and depression, work absence, and reduced quality of life (Kalameri et al 2009).

These findings support the hypothesis that in patients with MSK pain, generic factors including demographics, pain characteristics, psychological or social factors may be more important in the prediction of future outcome (prognosis) than the specific pain site or the assumed cause of pain or diagnosis. Prognosis rather than diagnosis may provide a framework for clinical practice, integrating biological, psychological and social information to support more effective and efficient care (Croft et al 2015). This may generate opportunities for the design and evaluation of interventions that target potentially modifiable prognostic factors regardless of pain site or diagnosis. This study tested this hypothesis by investigating the course of pain and limitations in function in trial participants with MSK pain in different sites, identify generic prognostic factors, and explore to what extent the association between prognostic factors and outcome in trial participants is modified by pain site. Trial data where treatment was randomly allocated was used to reduce the risk of treatment bias in this prognostic study.

METHODS

Design and Study Participants

This study was based on secondary analysis of individual patient data from seven randomised clinical trials carried out within the Arthritis Research UK Primary Care Centre (Keele University, UK), investigating a range of interventions with patients recruited from general practice, and published between 2004 and 2011. Information specific to the trials included in the present study is given below, and additional details are reported in Appendix 1.

Treatment Options for Pain in the Knee (TOPIK) trial: The aim of this three arm trial was to investigate the effects of physiotherapist-led advice and exercise (*intervention*) and enhanced pharmacy review (*intervention*) compared to an advice and exercise leaflet (*control*) in people aged 55 and over presenting in general practice with pain and/or stiffness lasting more than three months in one or both knees who were followed up at three, six and 12 months. Both interventions showed significantly larger short-term improvements in pain and function compared to control (Hay et al 2006).

Acupuncture Physiotherapy and Exercise (APEX) trial: The aim of this trial was to investigate whether adding acupuncture to a course of physiotherapist-led advice and exercise leads to greater pain relief in patients with knee pain (of any duration) attributable to OA (Hay et al 2004). The APEX trial included three arms, all delivered by physiotherapists: a course of advice and exercise (*control*), advice and exercise with true acupuncture (*true)* and advice and exercise with sham acupuncture. Participants were followed up at six weeks, six months and 12 months. The primary outcomes showed no additional benefit for true compared to sham or control (Foster et al 2007).

Low Back Pain (LBP) Trial: This trial compared the effects of a course of traditional physiotherapist-led management including exercise and manual therapy (*intervention*) with those of a brief course of pain management provided by physiotherapists (*intervention*) in patients with non-specific LBP (of less than three months duration). Participants were followed up at three and 12 months with the results indicating no significant or clinically important differences between the two treatments (Hay et al 2005).

Screening and Targeted Treatment for Back Pain (STarT Back) trial: This trial compared a stratified care approach consisting of prognostic stratification into low, medium and high risk subgroups followed by targeted treatment (*intervention)* with current best care (*control*) for patients presenting with non-specific LBP (of any duration). Changes in back pain-related disability and pain catastrophising were investigated at four and 12 months follow-up. The findings showed significantly superior pain and function outcomes in the stratified care arm compared to control (Hill et al 2011).

Physiotherapy Arthritis Research UK Neck Trial Hands-on and Electrotherapy Research (PANTHER): PANTHER investigated the outcomes of physiotherapist-led advice and exercise in addition to manual therapy (*intervention*) and advice and exercise with pulsed shortwave diathermy (PSWD) (*intervention*) compared to advice and exercise alone (*control*) in patients with neck pain (of more than four weeks duration). Outcomes were assessed at six weeks and six months. No significant differences were found between the treatment arms (Dziedzic et al 2005).

SPIRIT (Shoulder Physiotherapy and Injection RandomIsation Trial): This trial in patients with unilateral shoulder pain (of more than four weeks duration) compared outcomes of pain and function between participants randomly allocated to either a course of physiotherapist led advice, manual therapy and ultrasound (as required) (*intervention*) or corticosteroid injection (*intervention*). Participants were followed up at six weeks, six months and 18 months. The results showed no significant differences in outcomes across the treatment arms (Hay et al 2003).

Tennis Elbow trial: The objective of this three arm trial was to compare the effects of corticosteroid injection (*active*), Non-Steroidal Anti-Inflammatory Drugs (NSAIDs, *intervention*) and unmarked vitamin C (*placebo*) in patients with elbow pain (of any duration) attributable to lateral epicondylitis (tennis elbow) at four weeks, six months and 12 months follow-up. The results showed better outcomes for corticosteroid injection at four weeks, however more participants in this group showed relapse at long-term follow-up (Hay et al 1999).

Outcome measures

The outcome measures for all analyses were changes in (1) pain intensity and (2) limitation in function. Although most trials used a 0-10 point visual analogue scale (VAS) or numerical rating scale (NRS) to assess pain intensity, the trials used different pain-specific instruments to measure functional limitation. This outcome was measured with the Western Ontario and McMaster Universities Arthritis Index (WOMAC) function score (Bellamy et al 1988) in the knee pain trials (TOPIK, APEX); Roland Morris Disability Questionnaire (RMDQ) (Roland & Morris 1983) in the LBP trials (STarT Back, LBP trial); Northwick Park Neck Pain Questionnaire (NPPQ) (Leak et al 1994) in the neck pain trial (PANTHER); Shoulder Disability Questionnaire (SDQ) (Croft et al 1994) in the shoulder pain trial (SPIRIT), and a 1-10 NRS in the Tennis Elbow trial. In all datasets the scores for pain and functional limitation were transformed onto a 0-10 scale with higher scores indicating more severe pain or functional limitation, to allow comparison of descriptive results across studies and pooling of data for analysis. Evaluation of pain and function outcomes was based on change from baseline, and hence higher values denote greater improvement in pain/functional limitation.

Prognostic factors

Potential baseline prognostic factors, that were expected to be associated with outcomes of pain and disability based on existing evidence, were identified from the available variables within each dataset. Variables were subsequently recoded where needed to ensure consistency between datasets. Variables included for analysis were pain score (0-10), function score (0-10), age (continuous scale), gender, duration of pain episode (less than 1 month, 1-3 months, 3+ months), manual work (Manual versus Non-Manual occupation), presence of widespread pain according to American College of Rheumatology criteria (Wolfe et al 1990) (yes/no), presence of multisite pain if pain at more than one site was reported, and mood problems (none, moderate, extremely anxious or depressed, using the anxiety/depression item from EQ5D-3L) (EuroQol Group 1990). The Tennis Elbow trial did not collect information on anxiety and depression, and therefore this dataset was excluded from analyses of prognostic factors.

Analysis

Clinical course: To ensure comparable follow-up points, short-term and long-term time points were identified for each trial, selecting the scores nearest to three months follow-up as the short-term time point, and the scores at 12 months or later as the long-term time point. PANTHER did not include scores after six months follow-up and was therefore only included in the analysis of short-term outcomes. The baseline and follow-up outcome scores data from each of the seven trials were then merged. Pain duration did differ between the trials, partly as a result of the different study inclusion criteria; three studies (APEX, PANTHER and SPIRIT) did not include patients with pain of less than four week's duration, and the LBP trial did not include patients with pain of more than three months duration. In order to describe the course of symptoms within each trial overall, outcome scores for pain and function limitation at each follow up point were presented in a graph (Figures 1 and 2).

Changes from baseline at short- and long-term follow-up were calculated for both pain intensity and functional limitation. Linear regression was used to analyse changes in pain and functional limitation. Univariable models were computed with pain site (back, knee, shoulder, neck, and elbow) as the determinant in order to investigate differences in outcome for these separate pain sites, whilst adjusting for baseline pain and functional limitation only (as baseline levels are often found to be the strongest predictors of future pain and disability (e.g. Bot et al 2005; van der Waal et al 2005; Campbell et al 2013; Gustavsson et al 2013)).

Prognostic factors: Additional potential prognostic factors (all listed in Prognostic factors above) were then included to investigate which variables predicted outcome regardless of pain site or treatment. Analysis of the effectiveness of specific interventions was not an objective of this study, but intervention was included in the analysis as a potential confounder. Interventions were broadly classified into intervention, sham/placebo, or control depending on the nature of the treatment, where participants allocated to control or sham interventions often continued to receive care as usual (see trial descriptions above). The use of data from RCTs with random allocation of participants reduced the risk of treatment bias (also referred to as the treatment paradox), which may occur in prognosis studies using observational data if individuals with more severe or complex disease are more likely to receive more intensive treatment. If such treatment is effective, it will influence prognosis, and thereby also the association between potential prognostic factors and outcome (Schuit et al. 2013). Collinearity was examined by computing a variance inflation factor (VIF) for each factor in the model, where a value greater than 10 would indicate potential collinearity.

As data were clustered within trials, further analysis (with all the same prognostic factors included in the previous analysis) explored variation in outcomes across pain sites taking into account clustering of data within trials using random effects modelling: firstly through random intercept models, and secondly by random slope models to investigate the potential effect of clustering within pain sites. Comparison between models (random intercept and random slope) was based on a Likelihood-Ratio Test (LRT), where a significant difference between the two models implied that a random slope model is preferred. Intraclass Correlation Coefficients (ICCs) were estimated for each model to represent the correlation of the outcome within trials. A high ICC would imply that outcome scores were highly dependent on trial identification and analysis would require random effect constraints.

Moderation by pain site: Interaction terms (prognostic factor*pain site) for all prognostic factors with all pain sites were added to the random effects model (in addition to all the prognostic factors previously included). In order to explore to what extent the strength of associations between prognostic factors and outcomes varied across pain sites, these interactions of all prognostic factors with all pain sites were added to the random effects model one at a time, and in turn (replacing the previous interaction).

RESULTS

Participants

The number of participants from each included trial ranged from 164 to 851, resulting in a total sample of 2,651 participants, with individual participants' data for 2,483 available for analysis. Table 1 presents baseline characteristics and baseline scores of outcome measures for each of the trials. The mean age of participants ranged from 41 to 68 years, and 47% to 64% were female. Widespread pain was reported by between 5% and 39% of participants, with the median number of pain sites (potential scale: 0-49) varying between 3 and 8 sites across trials. Pain duration varied widely in the trial populations, with the APEX trial only including patients with pain of more than three months duration, the LBP trial focusing on patients with pain of less than three months duration, and the remaining trials including between 29% and 86% of patients who reported pain lasting for more than three months.

Course of pain intensity and function

Mean scores over time for both pain intensity (Figure 1) and functional limitation (Figure 2) show considerable variation between trials in terms of short-term improvement in these outcomes, yet in all trials most improvement of symptoms occurred over the first three months, with little further change over the subsequent 3 to 18 months of follow-up.

Analysis of changes in pain and function adjusted for baseline scores portrayed different relationships depending on pain site (Table 2). Participants in the LBP trials showed larger short-term improvements in both pain and function compared to those with pain at other body sites, although the difference was small, and not statistically significant for pain intensity in LBP participants compared to those with shoulder pain. Using LBP as the reference, participants of the shoulder or elbow trials showed larger improvements for both pain intensity and function at long-term follow-up, whereas those with knee pain showed less improvement in pain and function. The mean pain intensity score at 12 months (adjusted for differences in baseline values) was almost 2 points higher for participants with knee pain compared to LBP.

Prognostic factors

Increasing age, longer pain duration, manual work, presence of widespread pain, and mood problems (moderate/extreme anxiety or depressive symptoms) were significantly associated with poor outcome (smaller change in both pain and function), regardless of pain site and adjusted for intervention classification for both follow-up time-points, short and long term, with adjusted mean

differences ranging between 0.3 and 1.7 (Table 3). Higher levels of functional limitation at baseline were associated with larger improvements in function (adjusted mean difference 0.54 and 0.56, short- and long-term follow-up, respectively), but appeared to indicate slightly poorer pain outcomes (-0.16 and -0.18, respectively). A similar effect was seen for higher baseline levels of pain, although the impact on change in function was smaller. Females appeared to have larger improvements in long-term function compared to males, adjusted for other potential confounders (adjusted mean difference 0.25, 95% CI 0.05, 0.45). There was no evidence of collinearity in either model, with all VIF values for each confounder less than 2.10.

The random effects model indicated that ICCs were small, and the model confirmed significant variation in outcomes between trials for outcomes of short-term pain intensity (p=0.026) (ICC=0.009), long-term pain intensity (p=0.018) (ICC=0.013) but not for short-term functional limitation (p=0.062) (ICC=0.008), and variation was not significant for long-term functional limitation (p>0.9) (ICC<0.001). Fitting a random slope for pain site showed no significant improvement of the models, indicating that there was no significant variation in outcomes across individual trials (other than that explained by the fixed effects) i.e. no relevant influence of clustering on outcome trajectories (all p-values >0.50), and therefore random effects at trial level was deemed sufficient for further analysis.

Variation in prognostic factors across pain sites

Significant interactions mostly concerned site of pain at the knee (Table 4). Participants with knee pain showed stronger associations with poor outcome for baseline levels of pain and function, increasing age, longer pain duration, and manual work, although the interaction of pain site with manual work was only significant for pain intensity. Few other interactions were found, although in participants with shoulder pain, increasing age and male gender were more strongly associated with poorer outcomes of pain and/or function compared to LBP (reference category). LBP was used as the reference category because of the extensive evidence base regarding its course and prognosis. No significant interactions were found for widespread pain and mood problems, indicating that these variables had a similar effect on outcome across all pain sites.

DISCUSSION

This study, using individual participant data from seven randomised clinical trials, showed a similar pattern of improvement in pain intensity and functional limitation outcomes regardless of the site of MSK pain. An improvement in mean pain intensity and function scores was observed over the initial

three-month post-randomisation follow-up period of the trials, after which little further change occurred. Despite this similar overall pattern there were significant differences between pain sites in terms of the magnitude of improvement at short- and long-term follow-up. Participants with LBP showed the largest short-term improvements, whereas those with upper limb pain (shoulder or elbow) showed better long-term outcomes compared to other pain sites. Participants with knee pain (which were all older adults with pain attributable to OA) showed the least improvement during follow-up. In other words, improvement is seen across all pain sites but the magnitude of this improvement varies according to pain site. Increasing age, manual work, longer pain duration, mood, and presence of widespread pain were significantly associated with poor outcome, but testing for interactions showed some variation between pain sites in the strength of associations, particularly for knee pain where increased age, manual work and increased pain duration were stronger predictors of poor outcome compared to other sites.

Patterns of improvement

Our findings are consistent with a previous systematic review which investigated the pattern of symptom improvement in patients receiving different primary care treatments for non-specific LBP (Artus et al 2010). The review found that across a large number of trials and different types of treatment, a similar pattern of improvement emerged; rapid improvement within six weeks, followed by a slower improvement up to six months post-randomisation. The review authors proposed several reasons for these findings, including natural history of LBP, regression to the mean in people seeking care when pain levels are high, and the potential influence of variables other than specific treatment effect, including patient characteristics (prognostic factors) or therapist effects. Our finding that higher baseline pain scores were associated with larger improvements in pain despite poorer follow-up function scores (and vice versa), indicates there is room for improvement and a potential role of regression to the mean in those with high baseline scores for pain or functional limitation.

Although the patterns of pain was similar across pain sites, the present study shows variation in the magnitude of improvement across pain sites. The largest differences were found for participants of the two knee pain trials who showed poorer outcomes, and stronger associations between some prognostic factors (duration, age, manual work) and future outcome. However, these trials included people with knee pain attributed to osteoarthritis, reflecting a presentation of pain that is more likely to be characterised by persistent or recurrent pain and function over long periods of time, or simply by characteristics of the sample such as older age and longer duration of pain. Cohort studies

investigating long-term (5-7 year) trajectories of pain and function in people with knee OA have identified distinct subgroups with varying long-term symptom trajectories, often strongly associated with baseline levels of pain and function. These trajectories were classified as improving in 3-12% and as persistent-mild in 28-35% of participants, with other subgroups (40-60%) showing moderate to severe symptoms over long periods of time (Collins et al 2014; Nicholls et al 2014; White et al 2016). This confirms the more persistent course of pain and function in knee OA populations compared to other musculoskeletal pain presentations included in our analysis.

Generic prognostic factors

We did not specifically include trials focusing on people with pain at multiple sites, as we aimed to test the hypothesis that factors may predict outcome regardless of the site of pain. However, consistent with previous findings from observational studies, a significant proportion of participants did have widespread pain: approximately 25% of all trial participants met ACR criteria for widespread pain, mainly those with knee pain or LBP. This may have influenced our results regarding the interaction between pain site and prognostic factors, but also highlights the importance of assessing and investigating more generic aspects of pain presentation and not focusing on the regional pain site only. Widespread pain was included in the analyses as a potential prognostic factor, so the results reflect the impact of this factor.

The results from our study suggest that clinical decisions regarding treatment should therefore not be based on the site of MSK pain only. Croft et al (2015) have recently summarised evidence for a prognosis-based rather than a diagnosis-based framework for clinical decision-making, on the basis that the former provides a more biopsychosocial perspective and is perhaps more useful in presentations which have a less definitive biomedical diagnosis, as is the case for many patients with MSK pain. Studies that have investigated the predictive value of diagnostic information in regional pain presentations have not found diagnosis to be a strong prognostic factor (Spies-Dorgelo et al 2008; Chester et al 2016). These findings as well as the results from our study suggest that it is important to shift attention towards prognostic evidence when in the management of musculoskeletal conditions, rather than focusing on pain site and diagnosis only.

Recent evidence shows that subgrouping LBP patients based on risk of persistent disabling pain and matching the subgroups to different treatments is clinically and cost-effective (Hill et al 2011; Foster et al 2014). The results of this study add to that evidence, and together they highlight the need to develop and test approaches that subgroup MSK pain patients based on their prognosis, and then

match them to appropriate treatments. Prospective cohort studies can be used to derive and validate prognostic factors or multi-dimensional prognostic models across musculoskeletal pain presentations (e.g. Campbell et al 2016), but randomised clinical trials of sufficient size are needed to test if prognostic factors can predict a differential response to treatment, and to investigate the clinical and cost-effectiveness of stratified care approaches in which prognostic stratification is an important driver of treatment selection (Hingorani et al 2013; van der Windt & Dunn 2013; Croft et al 2015).

Strengths & Limitations

The results from this study are based on trial data collected between 1995 and 2008. It could be that data from more recent cohorts is different from that analysed here. However, the Artus et al (2014) review which included more recent data, from cohorts as well as trials, included age ranges and proportions of females that were within the ranges reported in the trials included in the present study. The inclusion criteria, presented in Appendix 1, show to what extent the results can be generalised to patients presenting to primary care with pain in various musculoskeletal pain sites. Each trial reflects a different target population, but the point of the present study was to investigate the predictive value of factors across these different populations, and also across the variations in pain and function scores between the different populations.

Differences in prognostic variables were difficult to resolve, and may have resulted in misclassification of exposures (information bias). More precise and consistent assessment of prognostic factors could have resulted in more precise estimates of associations. Our study may have underestimated the strength of association, although we have no reason to believe that misclassification would have resulted in a different direction of effect. This analysis looked at prognostic factors in trial participants (adjusting for intervention), but has not investigated predictors of differential treatment response (effect modification). The findings may therefore provide information to identify which patients may require further treatment, but not which specific treatments may be most beneficial to them. This paper also presents only a small number of trials with limited data on prognostic factors, and lack of consistency amongst the measures used to assess the prognostic factors and outcomes. Although prognostic factors were *a priori* selected based on existing evidence, the analysis did include a large number of interaction tests, given the different outcomes, time points and prognostic factors, which could have resulted in spurious

findings. However, the use of IPD from multiple trials is important to obtain a sufficiently large sample size to test for *a priori*-defined interactions (Debray et al 2015), and results were fairly consistent across trials.

Finally, while we considered standardising the function scales used as outcome measures we instead chose to transform them. As the scales differed across studies, standardisation may have allowed for more comparability. However, the distributions of scores across each of the function scales were similar (see Table 1) and allowed easier interpretation of the regression coefficients. To our knowledge, few studies have assessed and compared prognostic factors across different pain sites, in particular using trial data – where effects of treatment are less likely to influence associations between prognostic factors and outcome. Our study has added evidence regarding factors that predict outcome across different pain sites, which may inform the care for patients with various musculoskeletal conditions. The results of this exploratory analysis should be interpreted with caution, but they support further investigation of the predictive value and clinical utility of generic prognostic factors in patients with MSK pain across a range of pain sites.

Implications

This study identified two potentially modifiable factors (manual work and mood) that could be targeted during treatment. Mood has been found to be modifiable in physiotherapy settings where physiotherapists have been trained to target this factor (e.g. Lamb et al 2010; Hill et al 2011), and subsequent mediation analyses carried out on both of these trials to investigate how the interventions worked identified change in mood as a mediator, further highlighting its importance as a treatment target (Mansell et al 2016; Fordham et al 2016).

While work may be assumed to be a more difficult factor to address in clinical practice, a previous trial which provided vocational advice to those who were off sick due to LBP was found to successfully reduce the number of sick days taken (Wynne-Jones et al 2017). The intervention enabled GPs and nurses to refer patients to Vocational Advisors who could assist patients with obstacles to returning to work.

Conclusions

Analysis of individual patient data from multiple trials confirms the role of baseline levels of pain and function, widespread pain and mood problems as consistent predictors of poor outcome, providing

evidence to support prognostic stratification based on these factors, and offering opportunities for future investigation of the effectiveness of targeted treatment approaches across pain presentations.

Acknowledgements

The project was supported by a grant (project no. 84) from the National School for Primary Care (NIHR) School for Primary Care Research. DJG was funded by a NIHR School for Primary Care Research Doctoral Training Studentship.

KD is part funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Research and Care West Midlands and by a Knowledge Mobilisation Research Fellowship (KMRF-2014-03-002) from the NIHR.

NF, an NIHR Senior Investigator, is supported through an NIHR Research Professorship (NIHR-RP-011-015).

DvdW is a member of PROGRESS Medical Research Council Prognosis Research Strategy (PROGRESS) Partnership (G0902393/99558).

GM is supported by NIHR School for Primary Care Seedcorn Funding.

EH is a Senior NIHR Investigator.

The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

The authors have no conflicts of interest to declare.

Danielle van der Windt and Daniel Green came up with the original idea for the paper; Daniel Green carried out the statistical analysis and drafted the original paper; and Martyn Lewis, Gemma Mansell, Majid Artus, Krysia Dziedzic, Elaine Hay and Nadine Foster all provided interpretation of the results and helped to revise the original manuscript draft. All authors (Daniel Green, Martyn Lewis, Gemma Mansell, Majid Artus, Krysia Dziedzic, Elaine Hay, Nadine Foster, and Danielle van der Windt) have read and approved the paper.

References

Artus, M., van der Windt, D.A., Jordan, K.P., Hay, E.M. (2010). Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: A systematic review of randomized clinical trials. Rheumatology (Oxford) 49(12),2346-2356.

Artus, A., van der Windt, D., Jordan, K.P., Croft, P.R. (2014). The clinical course of low back pain: A meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. BMC Musculoskelet Disord 15,68.

Artus, M., Campbell, P., Mallen, C.D., Dunn, K.M., van der Windt, D.A.W. (2017). Generic prognostic factors for musculoskeletal pain in primary care: A systematic review. BMJ Open, 7,e012901.

Bastick, A.N., Runhaar, J., Belo, J.N., Bierma-Zeinstra, S.M. (2015). Prognostic factors for progression of clinical osteoarthritis of the knee: A systematic review of observational studies. Arthritis Res Ther 17,152.

Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt, L.W. (1988). Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 15(12), 1833-1840.

Bot, S.D., Van der Waal, J.M., Terwee, C.B., van der Windt, D.A., Scholten, R.J., Bouter, L.M., Dekker, J. (2005). Predictors of outcome in neck and shoulder symptoms: A cohort study in general practice. Spine 30(16),E459-E470.

Bruls, V.E., Bastiaenen, C.H., de Bie, R.A. (2015). Prognostic factors of complaints of arm, neck, and/or shoulder: A systematic review of prospective cohort studies. Pain 156(5),766-788.

Campbell, P., Foster, N.E., Thomas, E., Dunn, K.M. (2013). Prognostic indicators of low back pain in primary care: Five-year prospective study. J Pain 14(8),873-883.

Campbell, P., Hill, J.C., Protheroe, J., Afolabi, E.K., Lewis, M., et al. (2016). Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. J Pain Res 9,807-818.

Carnes, D., Parsons, S., Ashby, D., Breen, A., Foster, N.E., Pincus, T., Vogel, S., Underwood, M. (2007). Chronic musculoskeletal pain rarely presents in a single body site: Results from a UK population study. Rheumatology (Oxford) 46(7),1168-1170.

Chester, R., Jerosch-Herold, C., Lewis, J., Shepstone, L. (2016). Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: A multicentre longitudinal cohort study. Br J Sports Med. doi: 10.1136/bjsports-2016-096084.

Chou, R., Shekelle, P. (2010). Will this patient develop persistent disabling low back pain? JAMA 303(13),1295-1302.

Collins, J.E., Katz, J.N., Dervan, E.E., Losina, E. (2014). Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 22(5),622-630.

Croft, P., Pope, D., Zonca, M., O'Neill, T., Silman, A. (1994). Measurement of shoulder related disability: Results of a validation study. Ann Rheum Dis 53(8),525-528.

Croft, P., Altman, D.G., Deeks, J.J., Dunn, K.M., Hay, A.D., et al. (2015). The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. BMC Med 13(1),20.

Debray, T.P.A., Moons, K.G.M., van Walkenhof, G., Efthimiou, O., Hummel, N., Groenwold, R.H.H., Reitsma, J.B. (2015). GetReal Methods Review Group. Get real in individual participant data (IPD) meta-analysis: A review of the methodology. Research Synthesis Methods 6(4),293-309.

De Vos Andersen, N-B., Kent, P., Hjort, J., Høyrup Christiansen, D. (2017). Clinical course and prognosis of musculoskeletal pain in patients referred for physiotherapy: Does pain site matter? BMC Musculoskelet Disord 18,130.

The EuroQol Group. (1990). EuroQol-a new facility for the measurement of health-related quality of life. Health Policy 16(3),199-208.

Dziedzic, K., Hill, J., Lewis, M., Sim, J., Daniels, J., Hay, E.M. (2005). Effectiveness of manual therapy or pulsed shortwave diathermy in addition to advice and exercise for neck disorders: A pragmatic randomized controlled trial in physical therapy clinics. Arthritis Rheum 53(2),214-222.

Fordham, B.A., Lamb, S.E., Hansen, Z., Lall, R. & Ji, C. (2016). Explaining how cognitive behavioural approaches work for low back pain: Mediation analysis of the Back Skills Training Trial (BeST). Presented at the Low Back Pain Forum 2016.

Foster, N.E., Thomas, E., Barlas, P., Hill, J.C., Young, J., Mason, E., Hay, E.M. (2007). Acupuncture as an adjunct to exercise based physiotherapy for osteoarthritis of the knee: Randomised controlled trial. BMJ 335,436.

Foster, N.E., Mullis, R., Hill, J.C., Lewis, M., Whitehurst, D.G.T., et al. (2014). Effect of stratified care for low back pain in family practice (IMPaCT Back): A prospective population-based sequential comparison. Ann Fam Med 12(2),102-111.

Gustavsson, C., Bergstrom, J., Denison, E., von Koch, L. (2013). Predictive factors for disability outcome at twenty weeks and two years following a pain self-management group intervention in patients with persistent neck pain in primary health care. J Rehabil Med 45(2),170-176.

Hartvigsen, J., Natvig, B., Ferreira, M. Is it all about a pain in the back? (2013). Best Pract Res Clin Rheumatol 27(5),613-623.

Hay, E., Paterson, S., Lewis, M., Hosie, G., Croft, P. (1999). Pragmatic randomised controlled trial of local corticosteroid injection and naproxen for treatment of lateral epicondylitis in primary care. BMJ 319(7215),964-968.

Hay, E.M., Thomas, E., Paterson, S.M., Dziedzic, K., Croft, P.R. (2003). A pragmatic randomised controlled trial of local corticosteroid injection and physiotherapy for the treatment of new episodes of unilateral shoulder pain in primary care. Ann Rheum Dis 62(5),394-399.

Hay, E., Barlas, P., Foster, N., Hill, J., Thomas, E., Young, J. (2004). Is acupuncture a useful adjunct to physiotherapy for older adults with knee pain? The “acupuncture, physiotherapy and exercise” (APEX) study [ISRCTN88597683]. BMC Musculoskelet Disord 5,31.

Hay, E.M., Foster, N.E., Thomas, E., Peat, G., Phelan, M., Yates, H.E., Blenkinsopp, A., Sim, J. (2006). Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: Pragmatic randomised trial. BMJ 333(7576),995-998.

Hay, E.M., Mullis, R., Lewis, M., Vohora, K., Main, C.J., et al. (2005). Comparison of physical treatment versus a brief pain-management programme for back pain in primary care: A randomised clinical trial in physiotherapy practice. Lancet 365(9476),2024-2030.

Henschke, N., Ostelo, R.W., Terwee, C.B., van der Windt, D.A. (2012). Identifying generic predictors of outcome in patients presenting to primary care with non-spinal musculoskeletal pain. Arthritis Care Res 64(8),1217-1224.

Hill, J.C., Whitehurst, D.G.T., Lewis, M., Bryan, S., Dunn, K.M., et al. (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial. Lancet 378(9892),1560-1571.

Hill, J.C., Afolabi, E.K., Lewis, M., Dunn, K.M., Roddy, E., van der Windt, D.A., Foster, N.E. (2016). Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data. BMJ Open 6(10),e012445.

Hingorani, A.D., van der Windt, D.A., Riley, R.D., Abrams, K., Moons, K.G.M., et al. (2013). Prognosis Research Strategy (PROGRESS) 4: Stratified medicine research. BMJ 346,e5793.

Kamaleri, Y., Natvig, B., Ihlebaek, C.M., Benth, J.S., Bruusgaard, D. (2008). Number of pain sites is associated with demographic, lifestyle, and health-related factors in the general population. Eur J Pain 12(6),742-748.

Kamaleri, Y., Natvig, B., Ihlebaek, C.M., Benth, J.S., Bruusgaard, D. (2009). Change in the number of musculoskeletal pain sites: A 14-year prospective study. Pain 141(1-2),25-30.

Kooijman, M.K., Barten, D-J.A., Swinkels, I.C.S., Kuijpers, T., de Bakker, D., Koes, B.W., Veenhof, C. (2015). Pain intensity, neck pain and longer duration of complaints predict poor outcome in patients with shoulder pain – A systematic review. BMC Musculoskelet Disord 16:288.

Lamb, S.E., Lall, R., Hansen, Z., Castelnuovo, E., Withers, E.J., Nichols, V., Griffiths, F., Potter, R., Szczepura, A., Underwood, M., BeST trial team. (2010). A multicentred randomised controlled trial of a primary care-based cognitive behavioural programme for low back pain. The Back Skills Training (BeST) trial. Health Technol Assess 14 (41),1-253.

Leak, A.M., Cooper, J., Dyer, S., Williams, K.A., Turner-Stokes, L., Frank, A.O. (1994). The Northwick Pain Neck Pain Questionnaire, devised to measure neck pain and disability. Br J Rheumatol 33(5),469-474.

Mallen, C.D., Thomas, E., Belcher, J., Rathod, T., Croft, P., Peat, G. (2013). Point-of-care prognosis for common musculoskeletal pain in older adults. JAMA Intern Med 173(12),1119-1125.

Mansell, G., Hill, J.C., Main, C., Vowles, K.E. & van der Windt, D. (2016). Exploring what factors mediate treatment effect: Example of the STarT Back study high-risk intervention. J Pain 17(11), 1237-1245.

Nicholls, E., Thomas, E., van der Windt, D.A., Croft, P.R., Peat, G. (2014). Pain trajectory groups in persons with, or at high risk of, knee osteoarthritis: findings from the Knee Clinical Assessment Study and the Osteoarthritis Initiative. Osteoarthritis Cartilage 22(12),2041-2050.

Roland, M., Morris, R. (1983). A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. Spine 8(2),141-144.

Schuit E, Groenwold RH, Harrell FE Jr, de Kort WL, Kwee A, Mol BW, Riley RD, Moons KG. Unexpected predictor–outcome associations in clinical prediction research: causes and solutions. CMAJ, July 9, 2013, 185(10)

Spies-Dorgelo, M.N., van der Windt, D.A.W.M., Prins, A.P.A., Dziedzic, K.S., van der Horst, H.E. (2008). Clinical course and prognosis of hand and wrist problems in primary care. Arthritis Rheum 59(9),1349-1357.

Thomas, E., Dunn, K.M., Mallen, C., Peat. G. (2008). A prognostic approach to defining chronic pain: application to knee pain in older adults. Pain 139(2),389-397.

Valentin, G.H., Pilegaard, M.S., Vaegter, H.B., Rosendal, M., Ørtenblad, L., Vaeggemose, U., Christensen, R. (2016). Prognostic factors for disability and sick leave in patients with subacute non-malignant pain: A systematic review of cohort studies. BMJ Open 6,e007616.

van der Waal. J.A., Bot, S.D.M., Terwee, C.B.; van der Windt, D.A., Scholten, R.J., Bouter, L.M., Dekker, J. (2005). The course and prognosis of knee complaints in general practice. Arthritis Rheum 53(6),920-930.

van der Windt, D.A., Dunn, K.M., Spies-Dorgelo, M.N., Mallen, C.D., Blankenstein, A.H., Stalman, W.A. (2008). Impact of physical symptoms on perceived health in the community. J Psychosom Res 64(3),265-274.

van der Windt, D.A. (2010). The symptom of pain in populations. In Chronic Pain Epidemiology. From aetiology to public health. P.R. Croft, F. Blyth, D. Van der Windt D, eds. (Oxford University Press) pp. 31-50.

van der Windt, D.A., Dunn, K.M. (2013). Low back pain research: Future directions. Best Pract Res Clin Rheumatol 27(5),699-708.

Von Korff, M., Miglioretti, D.L. (2005). A prognostic approach to defining chronic pain. Pain 117,304-313.

White, D.K., Neogi, T., Nguyen, U.S., Niu, J., Zhang, Y. (2016). Trajectories of functional decline in knee osteoarthritis: the Osteoarthritis Initiative. Rheumatology (Oxford) 55(5),801-808.

Wolfe, F., Smythe, H.A., Yunus, M.B., Bennett, R.M., Bombardier, C., et al. (1990). The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum 33(2),160-172.

Wynne-Jones, G., Artus, M., Bishop, A., Lawton, S.A., Lewis, M., Jowett, S., Kigozi, J., Main, C., Wathall, S., Burton, A.K., van der Windt, D.A., Hay, E.M., Foster, N.E., SWAP Study team. (2017). Effectiveness and costs of a vocational advice service to improve work outcomes in patients with musculoskeletal pain in primary care: A cluster randomised trial (SWAP trial ISRCTN 52269669). Pain, epub ahead of print doi: 10.1097/j.pain.0000000000001075.

Table 1: Baseline characteristics of participants of each of the seven trials

Table 2: Univariable associations between pain site and short and long-term outcomes of change in pain and limitation in function[#]

Table 3: Multivariable associations (random effects linear regression, accounting for clustering within trials) between prognostic factors and outcomes of change in pain and limitation in function[#]

Table 4: Significant interactions of pain site with potential prognostic factors (each interaction term has been individually added to the multivariable random effects model (random intercept for trial), one at the time)

Figure 1. The course of pain intensity (mean scores, 0-10) for each trial

Figure 2. The course of limitation in function (mean scores, 0-10) for each trial

Table 1: Baseline characteristics of participants of each of the seven trials

		APEX (Knee)	**TOPIK (Knee)**	**LBP (Back)**	**STarT Back (Back)**	**PANTHER (Neck)**	**SPIRT (Shoulder)**	**Tennis Elbow Trial**	**Total**
Number of participants (n)		329	312	394	812	300	205	131	2483
Age (Mean, SD)		63.1 (8.7)	68.0 (8.1)	40.5 (11.7)	49.5 (14.4)	51.1 (13.8)	57.5 (13.4)	46.6 (7.9)	52.9 (15.0)
Female gender (n, (%))		198 (60.2)	200 (64.1)	208 (52.8)	469 (57.8)	185 (61.7)	110 (53.7)	61 (46.6)	1431 (57.6)
Duration (n, (%))	Less than 1 month	0 (0)*	11 (4.2)	340 (86.3)	142 (17.5)	14 (4.7)	75 (36.6)	26 (19.9)	608 (25.0)
	1-3 months	0 (0)*	27 (10.3)	53 (13.5)	186 (22.9)	57 (19.0)	70 (34.2)	61 (46.6)	454 (18.7)
	More than 3 months	329 (100)	224 (85.5)	1 (0.3)	484 (59.6)	229 (76.3)	60 (29.3)	44 (33.6)	1,371 (56.4)
Manual Occupation (n, (%))		163 (49.5)	175 (56.1)	241 (61.2)	433 (53.3)	146 (48.7)	111 (54.2)	75 (57.3)	1344 (54.1)
Widespread pain, ACR criteria (n, (%))		78 (23.7)	120 (38.5)	96 (24.4)	276 (34.0)	51 (17.0)	30 (14.6)	7 (5.3)	658 (26.5)
Baseline pain score (0-10 NRS)[$], Mean (SD)		5.83 (2.2)[#a]	5.92 (2.3)[#a]	5.56 (2.3)[£a]	4.86 (2.6)[£a]	4.89 (2.3)	5.09 (2.2)	5.21 (2.2)	5.28 (2.4)
Baseline function score (0-10 NRS)[$], Mean (SD)		4.42 (1.9)[#b]	4.39 (1.9)[#b]	5.63 (2.0)[£b]	4.03 (2.4)[£b]	3.61 (1.4)	4.70 (1.9)	3.61 (2.2)	4.36 (2.1)
Number of pain sites, Median (IQR)		6 (2.5-11)	8 (3-13)	8 (5-12)	n/a	n/a	4 (3-6)	3 (1-4)	6 (3-10)

Anxiety/ Depression (EQ5D item 5)	No	223 (68.2)	179 (60.3)	292 (74.3)	455 (56.6)	189 (63.2)	139 (68.8)	n/a	1477 (63.6)
	Moderate	94 (28.8)	114 (38.4)	95 (24.2)	307 (38.2)	107 (35.8)	57 (28.2)	n/a	774 (33.3)
	Severe	10 (3.1)	4 (1.4)	6 (1.5)	42 (5.2)	3 (1.0)	6 (3.0)	n/a	71 (3.1)
Follow-up times	1	6 weeks	3 months	3 months	4 months	6 weeks	6 weeks	4 weeks	
	2	6 months	6 months	12 months	12 months	6 months	6 months	6 months	
	3	12 months	12 months	n/a	n/a	n/a	18 months	12 months	

[*] APEX trial: duration of pain was measured as <1; 1-5; 5-10; >10 years. A clinical diagnosis of osteoarthritis was part of the eligibility criteria, hence a duration of more than three months was assumed for all. [#a]: Mean (SD) pain score for all knee participants= 5.88 (2.2); [#b]: Mean (SD) function score for all knee participants= 4.40 (1.9); [£a]: Mean (SD) function score for all back participants= 5.09 (2.5); [£b]: Mean (SD) function score for all back participants= 4.55 (2.4).

[$] Higher scores indicate higher levels of pain or function.

ACR = American College of Rheumatology criteria to classify presence of widespread pain; NRS= Numerical Rating Scale (on a scale of 0 to 10); IQR= Inter-Quartile Range; EQ5D = EuroQol, 5 Dimensional questionnaire; n/a = data not available (information on number of pain sites not collected in STarT Back/ PANTHER, EQ5D not collected in Tennis Elbow trial, no 3rd follow-up point in LBP, STarT Back or PANTHER trials)

Table 2: Univariable associations between pain site and short and long-term outcomes of change in pain and limitation in function[#]

		Pain (0-10)		Limitation in function (0-10)	
		Short term (≈3 months) Adjusted mean difference (95% CI)	**Long term (≥12 months)** Adjusted mean difference (95% CI)	**Short term (≈3 months)** Adjusted mean difference (95% CI)	**Long term (≥12 months)** Adjusted mean difference (95% CI)
Number of observations (n)		2,068	1,670	2,134	1,788
Pain site	Back (reference)	0	0	0	0
	Knee	-1.58 (-1.83, -1.34)***	-1.85 (-2.12, -1.57)***	-1.39 (-1.59, -1.19)***	-1.62 (-1.83, -1.40)***
	Neck	-1.08 (-1.40, -0.76)***	n/a	-0.95 (-1.22, -0.68)***	n/a
	Shoulder	-0.34 (-0.70, 0.03)	1.21 (0.78, 1.64)***	-1.15 (-1.46, -0.85)***	0.74 (0.40, 1.09)***
	Elbow	-0.58 (-1.02, -0.14)**	1.05 (0.58, 1.52)***	-0.43 (-0.80, -0.07)*	0.88 (0.49, 1.26)***

[#]Adjusted for baseline levels of pain and function
*p<0.05, **p<0.01, ***p<0.001
95% CI = 95% confidence interval; n= sample size used for each analysis.

Table 3: Multivariable associations (random effects linear regression, accounting for clustering within trials) between prognostic factors and outcomes of change in pain and limitation in function[#]

Potential prognostic factor		Pain (0-10)		Limitation in function (0-10)	
		Short term	**Long term**	**Short term**	**Long term**
Number of observations (n)		1,875	1,477	1,936	1,593
		Adjusted mean difference (95% CI)	Adjusted mean difference (95% CI)	Adjusted mean difference (95% CI)	Adjusted mean difference (95% CI)
Pain site	Back (ref)	0	0	0	0
	Knee	-0.98 (-1.30, -0.66)***	-1.07 (-1.45, -0.70)***	-0.77 (-1.04, -0.51)***	-0.74 (-1.03, -0.44)***
	Neck	-0.79 (-1.12, -0.47)***	n/a	-0.66 (-0.94, -0.39)***	n/a
	Shoulder	-0.37 (-0.73, -0.00)*	1.27 (0.83, 1.71)***	-1.19 (-1.50, -0.89)***	0.76 (0.41, 1.11)***
	Elbow	n/a	n/a	n/a	n/a
Baseline pain score (0-10 NRS)		0.68 (0.63, 0.73)***	0.74 (0.68, 0.80)***	-0.10 (-0.14, -0.05)***	-0.07 (-0.12, -0.02)**
Baseline function score (0-10)		-0.16 (-0.22, -0.10)***	-0.18 (-0.25, -0.11)***	0.54 (0.49, 0.59)***	0.56 (0.50, 0.62)***
Age (in years)		-0.02 (-0.03, -0.01)***	-0.02 (-0.03, -0.01)***	-0.02 (-0.02, -0.01)***	-0.02 (-0.03, -0.01)***
Gender	Male	0	0	0	0
	Female	0.04 (-0.17, 0.25)	0.20 (-0.05, 0.46)	-0.04 (-0.21, 0.13)	0.25 (0.05, 0.45)*
Duration	< 1 month	0	0	0	0
	1-3 months	-0.42 (-0.76, -0.08)*	-0.54 (-0.95, -0.14)**	-0.25 (-0.53, 0.02)	-0.29 (-0.61, 0.02)
	> 3 months	-1.08 (-1.37, -0.78)***	-1.31 (-1.66, -0.96)***	-0.75 (-0.99, -0.51)***	-1.03 (-1.30, -0.76)***

Manual work	Non-Manual	0	0	0	0
	Manual	-0.33 (-0.53, -0.12)***	-0.34 (-0.59, -0.08)**	-0.38 (-0.55, -0.21)***	-0.37 (-0.57, -0.17)***
Widespread Pain (ACR criteria)	No	0	0	0	0
	Yes	-0.39 (-0.62, -0.16)***	-0.55 (-0.83, -0.27)***	-0.39 (-0.58, -0.19)***	-0.56 (-0.78, -0.34)***
Mood: anxiety/ depression (EQ-5D, item 5)	No	0	0	0	0
	Moderate	-0.16 (-0.38, 0.06)	-0.30 (-0.56, -0.03)*	-0.11 (-0.29, 0.07)**	-0.37 (-0.58, -0.16)**
	Severe	-0.64 (-1.26, -0.03)*	-0.45 (-1.19, 0.29)	-1.04 (-1.53, -0.54)***	-0.54 (-1.11, 0.03)
Constant		1.06 (0.50, 1.61)	1.33 (0.64, 2.03)	1.59 (1.13, 2.04)	1.81 (1.27, 2.35)

 # All associations have also been adjusted for treatment category (intervention, sham/placebo, control), and every other variable listed in the table. n= sample size used for each analysis; 95% CI= 95% Confidence Interval; NRS= Numerical Rating Scale (on a scale of 0 to 10); ACR = American College of Rheumatology criteria to classify presence of widespread pain; EQ5D = EuroQol, 5 Dimensional questionnaire; n/a = data not available (EQ5D not collected in Tennis Elbow trial, therefore Tennis Elbow participants were not included in the multivariable analysis).

**p<0.05, **p<0.01, ***p<0.001*

Table 4: Significant interactions of pain site with potential prognostic factors (each interaction term has been individually added to the multivariable random effects model (random intercept for trial), one at the time)

Potential interactions		**Pain (0-10)**		**Limitation in function (0-10)**	
		Short term	**Long term**	**Short term**	**Long term**
		Coefficient (95% CI) [1]	Coefficient (95% CI)	Coefficient (95% CI)	Coefficient (95% CI)
Pain	Back	0	0	0	0
	Knee	-0.13 (-0.24, -0.03)*	-0.19 (-0.31, -0.07)**	-0.09 (-0.18, -0.00)*	-0.17 (-0.26, -0.07)**
	Neck	0.03 (-0.11, 0.16)	n/a	0.00 (-0.11, 0.11)	n/a
	Shoulder	0.15 (-0.01, 0.30)	0.17 (-0.01, 0.35)	0.08 (-0.05, 0.21)	0.11 (-0.03, 0.25)
Function	Back	0	0	0	0
	Knee	-0.24 (-0.36, -0.12)***	-0.21 (-0.35, -0.07)**	-0.26 (-0.36, -0.16)***	-0.26 (-0.37, -0.15)***
	Neck	-0.14 (-0.35, 0.07)	n/a	-0.17 (-0.35, 0.01)	n/a
	Shoulder	0.15 (-0.03, 0.32)	0.16 (-0.05, 0.36)	0.01 (-0.16, 0.13)	-0.02 (-0.19, 0.14)
Age	Back	0	0	0	0
	Knee	-0.02 (-0.05, 0.00)	-0.01 (-0.04, 0.01)	-0.02 (-0.04, 0.00)	-0.01 (-0.03, 0.01)
	Neck	0.02 (-0.00, 0.04)	n/a	0.01 (-0.01, 0.02)	n/a
	Shoulder	-0.02 (-0.05, -0.00)	0.01 (-0.03, 0.04)	-0.02 (-0.04, 0.00)	0.02 (-0.01, 0.04)
Manual work	Back	0	0	0	0
	Knee	-0.66 (-1.14, -0.18)**	-0.65 (-1.19, -0.10)*	-0.23 (-0.63, 0.17)	-0.31 (-0.74, 0.12)

	Neck	-0.32 (-0.93, 0.29)	n/a	0.04 (-0.47, 0.55)	n/a
	Shoulder	-0.13 (-0.83, 0.56)	0.07 (-0.76, 0.90)	-0.13 (-0.71, 0.45)	-0.10 (-0.77, 0.56)
Depression/ Anxiety	Back	0	0	0	0
	Knee (Mod)	-0.16 (-0.68, 0.35)	-0.03 (-0.62, 0.55)	-0.25 (-0.68, 0.18)	-0.22 (-0.69, 0.24)
	Knee (Sev)	-0.10 (-1.64, 1.43)	-0.11 (-1.83, 1.60)	-0.52 (-1.76, 0.71)	-0.71 (-2.08, 0.66)
	Neck (Mod)	0.71 (0.07, 1.36)*	n/a	0.25 (-0.29, 0.79)	n/a
	Neck (Sev)	-0.82 (-3.46, 1.82)	n/a	-1.62 (-3.83, 0.60)	n/a
	Shoulder (Mod)	0.05 (-0.72, 0.81)	0.30 (-0.64, 1.23)	-0.06 (-0.70, 0.58)	-0.02 (-0.77, 0.73)
	Shoulder (Sev)	-0.45 (-2.42, 1.52)	0.42 (-1.93, 2.78)	-1.20 (-2.85, 0.44)	-0.31 (-2.22, 1.60)
Duration	Back	0	0	0	0
	Knee (1-3)	-1.33 (-3.16, 0.50)	-1.20 (-3.44, 1.05)	-1.35 (-2.89, 0.18)	-1.57 (-3.19, 0.05)
	Knee (3+)	-1.03 (-2.64, 0.58)	-2.15 (-4.16, -0.15)*	-0.88 (-2.23, 0.47)	-1.58 (-3.00, -0.16)*
	Neck (1-3)	0.03 (-1.36, 1.42)	n/a	-0.11 (-1.27, 1.05)	n/a
	Neck (3+)	0.82 (-0.44, 2.09)	n/a	0.63 (-0.43, 1.69)	n/a
	Shoulder (1-3)	-0.60 (-1.46, 0.27)	-0.34 (-1.38, 0.69)	-0.55 (-1.27, 0.16)	0.38 (-0.44, 1.20)
	Shoulder (3+)	-0.11 (-0.99, 0.77)	0.08 (-0.97, 1.14)	-0.27 (-1.00, 0.46)	0.41 (-0.42, 1.24)
Gender	Back	0	0	0	0
	Knee	0.47 (-0.02, 0.96)	0.25 (-0.31, 0.81)	0.21 (-0.19, 0.62)	0.20 (-0.24, 0.64)
	Neck	0.08 (-0.55, 0.70)	n/a	-0.00 (-0.53, 0.52)	n/a

	Shoulder	0.62 (-0.07, 1.32)	0.62 (-0.21, 1.45)	0.31 (-0.27, 0.89)	0.76 (0.09, 1.43)*
Widespread pain	Back	0	0	0	0
	Knee	0.33 (-0.20, 0.85)	0.28 (-0.31, 0.87)	0.19 (-0.24, 0.62)	-0.01 (-0.47, 0.46)
	Neck	0.17 (-0.60, 0.94)	n/a	0.24 (-0.40, 0.89)	n/a
	Shoulder	0.05 (-0.92, 1.02)	0.03 (-1.11, 1.94)	-0.74 (-1.53, 0.05)	-0.33 (-1.24, 0.57)

**p<0.05, **p<0.01, ***p<0.001*

[1] *The coefficient reflects the additional effect on pain or functional limitations of both the predictor and pain in a specific site, above and beyond the combined effects of pain site and predictor alone (negative coefficient indicates that the interactions strengthens the combined effect on change in pain/function if the coefficient is negative in the previous table (Table 3) (similar for positive coefficients), however, contrasting coefficient (positive in Table 3 but now negative) indicates weakening of the combined effect)*

95% CI= 95% Confidence Interval; n/a= data not available (no longer term follow-up collected in PANTHER trial)

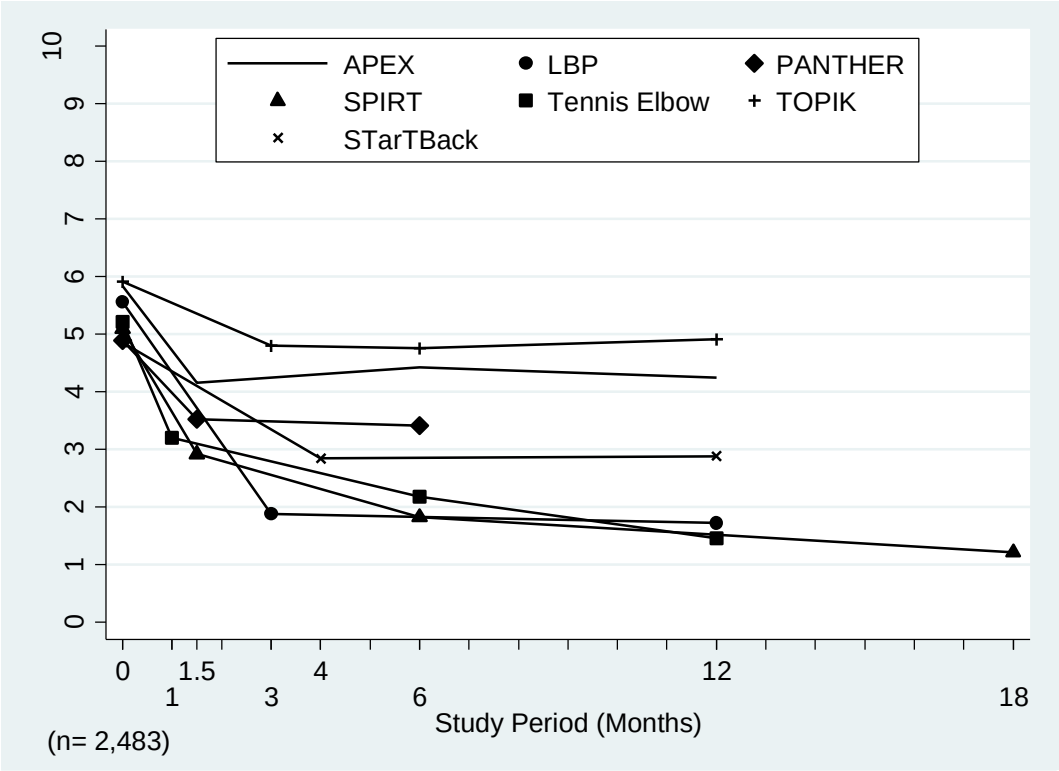

Figure 1. The course of pain intensity (mean scores, 0-10) for each trial

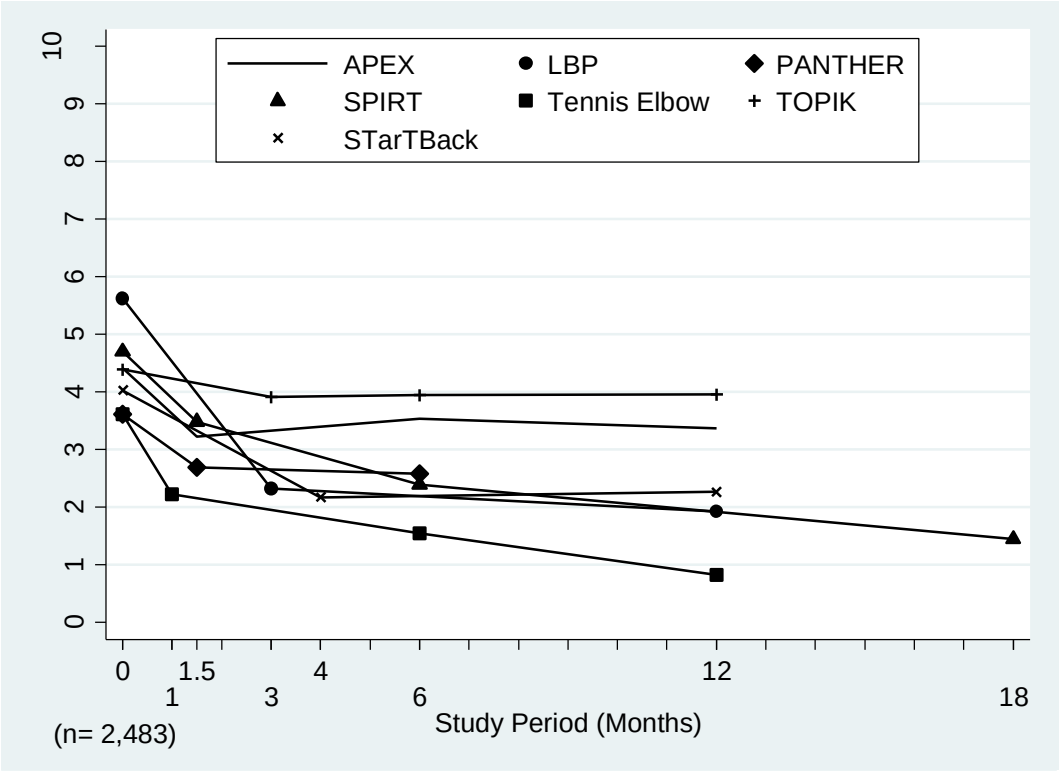

Figure 2. The course of limitation in function (mean scores, 0-10) for each trial

Appendix 1: Additional information about included studies

	APEX	TOPIK	LBP	StarT Back	PANTHER	SPIRIT	Tennis Elbow
Inclusion criteria	Adults aged 50y or older with knee pain and a clinical diagnosis of knee OA	Adults aged 55y and over consulting GP with knee pain, stiffness or both in one or both knees	Adults aged 18-64 years consulting their GP for the first or second time with non-specific LBP of less than 12 weeks duration	Adults aged at least 18 years with back pain of any duration, with or without associated radiculopathy	Adults aged 18 years or older with a clinical diagnosis of nonspecific neck pain (new episode) referred by physio by their GP	Adults aged 18 years and over consulting their GP with a new episode (no previous consultation in the last 12m) of unilateral shoulder pain	Adults aged 18-70 years who consulted GP with a new episode (no previous consultation in last 12m) of lateral epicondylitis
Time period of recruitment	November 2003-October 2005	May 2001-March 2004	July 2000-July 2002	June 2007-November 2008	June 2000-June 2002	June 1998-March 2000	November 1995-December 1997
Participation rate	GP referrals=1061; 352 randomised; 351 received allocated intervention	Retrospective record review + GP referrals = 691; 325 randomised; 311 received allocated intervention	544 assessed; 402 randomised; 315 received allocated intervention	1573 assessed; 851 randomised	735 assessed; 350 randomised; 332 received allocated intervention	237 referred; 207 randomised; 192 received treatment	182 referred; 164 randomised; 156 received treatment
Loss to follow-up	At 6w, 18 lost to follow-up; at 6m, 21 lost to follow-up; at 12m, 40 lost	At 3m, 44 lost to follow-up; 6m, 40 lost to follow-up; 12m, 53 lost to	At 3m, 83 lost to follow-up; at 12m, 73 lost to follow-up	At 4m, 162 lost to follow-up (withdrawals and non-responders); at 12, 40 lost to	At 6w, 29 lost to follow-up (non-responders) and 9 missing data on outcome	At 6w, 9 lost to follow-up; at 6m, 2 lost to follow-up	1 lost to follow-up at 12m in control arm

	to follow-up ITT analysis	follow-up ITT analysis	ITT analysis	follow-up (withdrawals and non-responders) ITT analysis	(Northwick Park); at 6m, 25 lost to follow-up (non-responders) and 10 missing data on outcome		