Physiotherapy 101 (2015) 243–251

Systematic review

Identifying potential moderators for response to treatment in low back pain: A systematic review

Tara Gurung [a], David R. Ellard [b], Dipesh Mistry [b], Shilpa Patel [b,*], Martin Underwood [b]

[a] *Warwick Evidence, Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry CV4 7AL, UK*
[b] *Clinical Trials Unit, Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry CV4 7AL, UK*

Abstract

Background Identifying which patients with non-specific low back pain are likely to gain the greatest benefit from different treatments is an important research priority. Few studies are large enough to produce data on sub-group effects from different treatments. Data from existing large studies may help identify potential moderators to use in future individual patient data meta-analyses.

Objective To systematically review papers of therapist delivered interventions for low back pain to identify potential moderators to inform an individual patient data meta-analysis.

Data sources We searched MEDLINE, EMBASE, Web of Science and Citation Index and Cochrane Register of Controlled Trials (CENTRALhttp://www.cochrane.org/editorial-and-publishing-policy-resource/cochrane-central-register-controlled-trials-central) for relevant papers.

Data extraction and data synthesis We screened for randomised controlled trials with ≥500 or more participants, and cohort studies of ≥1000 or more participants. We examined all publications related to these studies for any reported moderator analyses. Two reviewers independently did risk of bias assessment of main results and quality assessment of any moderator analyses.

Results We included four randomised trials ($n = 7208$). Potential moderators with strong evidence ($p < 0.05$) in one or more studies were age, employment status and type, back pain status, narcotic medication use, treatment expectations and education. Potential moderators with weaker evidence ($0.05 < p \leq 0.20$) included gender, psychological distress, pain/disability and quality of life.

Conclusion There are insufficient robust data on moderators to be useful in clinical practice. This review has identified some important potential moderators of treatment effect worthy of testing in future confirmatory analyses.

Keywords: Low back pain; Back pain; Randomised controlled trial; Cohort; Prospective studies

Background

Low back pain (LBP) is very common and has a large personal and societal cost [1]. Most LBP is classified as non-specific LBP (NSLBP) which affects one-third of the population each year [2]. There is good evidence to show that several treatment approaches are effective, and that some of these are cost-effective [2]. The effect sizes are of similar magnitude for different approaches [3–6]. However, the mean effect size from these treatments is, at best, small to moderate and may be short lived. Typically, the mean effect sizes, on current outcome measures, are substantially smaller than the minimally detectable change for an individual. Thus, most of the patients who receive a particular treatment will not gain a noticeable additional benefit from treatment [7]. At a population level, we have useful data on the management of LBP. What is not clear is how we can use these data to maximise the treatment benefit for the individual patient, or to identify those who will respond to different treatment and target treatment accordingly. Identifying which patients are

* Corresponding author. Tel.: +44 02476 150 405.
E-mail addresses: T.Gurung@warwick.ac.uk (T. Gurung), D.R.Ellard@warwick.ac.uk (D.R. Ellard), D.Mistry@warwick.ac.uk (D. Mistry), Shilpa.Patel@warwick.ac.uk (S. Patel), M.Underwood@warwick.ac.uk (M. Underwood).

http://dx.doi.org/10.1016/j.physio.2015.01.006

likely to gain the greatest benefit from different treatments for LBP is an identified research priority [8] and was one of the key recommendations for future research in UK National Institute for Health and Clinical Excellence (NICE) back pain guidelines [9].

In clinical practice, to try to maximise treatment benefit, subgrouping is used for patients with LBP despite lack of evidence that results vary between subgroups [10]. NICE considers identification of subgroups as an important part in their decision making on whether the technology is clinically effective or cost-effective [4]. In order to develop such subgroups a clear understanding of the potential moderators of treatment is required.

Many studies have examined predictors of outcome from LBP [11–13]. These do not, however, identify moderators; those factors indicating who is likely to gain largest benefit from a particular treatment. Mediators, measured during treatment, identify potential mechanisms that have an interactive effect on outcome [14]. This review solely focuses on moderators of treatment response; factors measured prior to randomisation that affect whether an individual has a greater, or lesser benefit from treatment [15]. Identification of potential effect modifiers needs sufficient statistical power to detect an interaction between the moderators and treatment [16].

Any RCT designed to test effects in subgroups will need to be several times larger than nearly all existing RCTs. Most trials simply compare the effects of two interventions with one primary outcome measure. More complex designs testing multiple baseline measures, and multiple interventions, would be implausibly large. However, many participants are now included in RCTs, in some cases testing similar interventions and most using very similar outcome measures. Combining data from these trials could provide a more cost-effective way of exploring and testing for moderator effects without the expense of a large costly and time consuming trial.

Aims & objectives

The aim of this systematic review was to inform hypothesis development for an individual patient data meta-analysis for moderators of therapist delivered interventions in RCTs. Therefore the question being addressed was are there subgroups of patients with low back pain, receiving therapist delivered interventions that do better or worse?

To achieve this our objectives were:

- To search the relevant literature in the field.
- To screen the literature based on predefined inclusion criteria.
- To extract data and quality assess the literature.
- To highlight the potential moderators from the literature to apply to an individual patient data meta-analysis.

Methods

Eligibility criteria

The following inclusion criteria was pre specified:

(a) RCTs with sample size of ≥500, non-RCTs and observational studies with sample size ≥1000 published in English language; see below for justification of the 500 cut-off.
(b) Participants aged 18 years or more with history of NSLBP of any duration.
(c) Therapist delivered interventions for LBP examining the effect of patient preference and expectations, and individual predictors.
(d) Primary and secondary analysis papers of RCTs seeking to identify predictors of response to treatment using a 'priori' and 'post hoc' subgroups and those looking for interaction between baseline variable and treatment.

We only included studies of people with NSLBP. We excluded studies with no comparison between two treatment groups and studies that did not report effect sizes for treatment by using moderator interactions.

Information sources

We searched MEDLINE (1948 to September 2011), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, EMBASE (1974 to September 2011), Web of Science and Citation Index and Cochrane Controlled Trial Registered (CENTRAL) databases for relevant papers. The searches were updated in May 2013 then again in July 2014.

Search criteria

Preliminary searches were carried out by using search terms such as 'low back pain' combined with keywords like 'subgroup', 'effect modifier' and 'moderator'. However this only yielded publications that had terms 'subgroup' in the title/abstract only, missing out publications that had the term 'subgroup' in the main text. We therefore re-ran searches using keywords ('trial') for RCTs and ('Observational', 'Cohort', 'Prospective studies') for non-RCTs or observational studies separately and then combining them with terms 'low back pain' (see Supplementary file 1). Hand searching and screening of included studies were carried out for additional studies.

Study selection and data extraction

Two authors (TG & DE) scanned titles and abstracts based on the pre-specified inclusion criteria. Data extraction was carried out by two reviewers (TG & DE) independently, using a standardised data extraction form. A third reviewer (MU) was available to consult if there were discrepancies.

A pre-piloted form was used to extract relevant data independently on study (e.g., author, country, design, sample size), participants (e.g., age, sex, level and year of education, employment status, back pain history and current episode of back pain), interventions (e.g., manipulation, exercise, cognitive behavioural therapy and acupuncture), and outcome characteristics (e.g., scale of measurement such as RMDQ and modified von Korff (MVK) scales of pain and disability). The extracted data were cross-checked by second reviewer and any disagreement were sorted by discussion.

One of the criteria for including clinical trials in this review is that the sample size was adequate for meaningful subgroup and/or interaction effect analyses. For the sample size criterion, we assumed that the outcome of interest is continuous and normally distributed, there are two treatment arms (intervention and control) and the potential moderator is binary. We used a simple model proposed by Lachenbruch [19] to determine the minimum sample size needed to test for an interaction effect. To test for a long-term (12 months) moderate standardised effect size of 0.5 for the interaction at a 0.05 level of significance and 80% power, a minimum data-set of 503 participants was needed. An additional file shows the sample size calculation in more detail [see Supplementary file 2]. Therefore any variables identified as moderators of treatment effect at $p < 0.05$ were credible for the purpose of our review. However, there might be a number of other variables whose effect approaches statistical significance and might be potential moderators that were not detected as the sample size of the trial is insufficient to make it a statistically significant finding.

To put this into context the effect sizes of the high-quality RCTs of therapist delivered interventions for low back pain are typically in range 0.12 to 0.23 [6]. Thus, any trial smaller than our size criterion would only be able to detect treatment moderation if the moderation effect was substantially larger than the main treatment effect.

Data items

For this study we did not pre-define the moderator variables of interest. We were, rather, seeking to identify the potential moderators identified by others to allow us to define a-priori the variables of interest in our subsequent IPD meta-analysis. Artificially restricting variables of interest at this stage of the process might have run the risk of introducing bias into the selection process.

Risk of bias

This was carried out by two reviewers (XX, XX) using the 'Cochrane Collaboration risk of bias tool'. The criteria used were: (a) method of randomisation, (b) allocation concealment, (c) incomplete outcome data, (d) selective outcome reporting, and (e) other source of bias [17]. Where a study had multiple publications, risk of bias assessment was conducted on the paper containing the main study findings.

Summary measures

The purpose of this review was to identify variables to be included in a subsequent individual patient data meta-analysis. For this reason we report all interactions with a p-value of ≤ 0.20 to ensure all possible moderators were identified and present the mean difference with 95% confidence interval for the interaction between treatment and baseline variable for each subgroup. Therefore we considered moderators as those with strong ($p < 0.05$) or weak evidence ($p < 0.20$, ≥ 0.05).

The primary outcome of clinical interest here is the size of the interaction between baseline variables and treatment. This is a measure of the differential sub-group effect. That is the size of the difference in the average effect of treatment between two groups defined by a baseline characteristic (moderator). Where provided in the original papers we report this as a point estimate with a 95% confidence interval. Where only a 'p' value is reported it is that which have presented. For our current purpose we are not actually trying to estimate the magnitude of any clinical effect we are simply seeking to identify potential moderators based on the level of significance of any interactions identified. It is thus the 'p' value that becomes our primary summary measure of outcome for this study.

Synthesis of results

Although the same moderators were investigated in several studies, it was not possible to perform meta-analyses due to statistical heterogeneity.

Quality of moderator analysis

The quality assessment of subgroup analyses within studies was carried out using the Pincus criteria [18] which classifies the level of evidence into confirmatory evidence or exploratory evidence. Members of the reviewing team who were authors on any included studies did not participate in the quality assessment exercises.

Results

We identified 7208 citations in total including all the updated searches. 6294 were removed based on title, abstract and duplicates. The full texts of 64 papers were retrieved for further evaluation; 60 of these did not meet our inclusion criteria (Fig. S1).

We included analyses from four RCTs [3,5,20,24] and their published secondary papers [7,20–24] ($n = 5514$), henceforth named the 'UK BEAM', 'BeST', 'Witt' and 'Cherkin' trial (Table 1). Our sample size calculation for our

Table 1
Characteristics of included studies.

Study ID	Inclusion criteria/sample size (*N*)	Intervention	Follow up	Outcomes
BeST Trial United Kingdom (Lamb et al., HTA, 2010, 14; Underwood et al., Arthritis Care Res, 2011;63:1271)	Inclusion criteria: Participants had attended general practice with sub-acute and chronic LBP, who were experiencing symptoms of at least moderate troublesomeness for >6 weeks. Low back pain presents in a wide variety of ways, and typically patients present with a spectrum of severity and chronicity, participants had to be aged 18 years or older, participants were able to give informed consent *N* = 701	Active management (advice only) Advice plus cognitive-behavioural therapy	12 Months	Roland Morris Disability Questionnaire (RMDQ) Modified von Korff (MVK) scales of pain and disability
Cherkin 2009 USA (Cherkin et al., Arch Intern Med, 2009;169:858; Underwood et al., Arthritis Care Res, 2011;63:1271)	Inclusion criteria: LBP as presenting symptom on the day of recruitment, written consent to participate in the study, and age above 19 years *N* = 638	Individualised treatment Standardised acupuncture Simulated acupuncture Usual care	8 Weeks 52 Weeks	Symptoms Bothersomeness score Back related dysfunction (Roland score)
UK BEAM Trial Kingdom (UK BEAM, BMJ, 2004;329:1377; Underwood et al., Rheumatology, 2007;46:1297)	Inclusion criteria: 18 and 65 years of age; consulted with simple low back pain -pain of musculoskeletal origin in the area bounded by the lowest palpable ribs, the gluteal folds, and the posterior axillary lines, including pain referred into the legs provided it was mainly above the knee; they had score of four or more on the Roland disability questionnaire at randomisation; agreed to avoid physical treatments, other than trial treatments, for three months; experienced pain every day for the 28 days before randomisation or for 21 out of the 28 days before randomisation; and 21 out of the 28 days before that *N* = 1334	Manipulation, exercise, Manipulation followed by exercise (combined treatment)	3 Months 12 Months	Scores on the Roland Morris disability questionnaires
Witt 2006 Germany (Witt et al., Clin J Pain, 2011;27:550)	Inclusion criteria: Clinical diagnosis of chronic low back pain with disease duration of more than 6 months; age ≥18 years; and provision of written informed consent *N* = 2841 (Randomised)	Immediate acupuncture treatment Delayed acupuncture after 3 months	6 Months	Back function or pain improvement scores

inclusion criteria of an overall sample size of ≥500 was based on a two arm trial i.e. 250 participants per arm. The Cherkin trial met the inclusion criteria for an overall sample size of ≥500, however this was a four arm trial meaning the number of participants in each arm was around 150. As this trial still generated useable information, we decided to include it and revisit our search results to identify any trials that had an overall sample size of ≥300. We did not identify any further trial meeting this revised criterion.

For the Witt trial the moderator analyses were included in the main paper [24]. For UK BEAM, BeST & Cherkin the moderator analyses were presented in a secondary paper [7,22,23]. UK BEAM and BeST were carried out in the UK, Witt in Germany and Cherkin in the USA. Sample sizes ranged from 600 to 2841. Mean age ranged from 47 to 53 years and the majority of the participants were female (56% to 62%). The interventions in the included studies are acupuncture [22,24], group cognitive behavioural approach (BeST) [21], group exercise (UK BEAM) [7], manual therapy (UK BEAM) and manual therapy followed by exercise (UK BEAM).

Outcomes reported in the studies were Roland Morris Disability Questionnaire (RMDQ), back-related dysfunction and bothersomeness score, MVK (Modified Von Korff) pain and disability, back function and pain improvement. Total follow-up duration and the unit of measure used (e.g. months or weeks) varied across the trials; ranging between three months (12 weeks) and 52 weeks. The characteristics of included studies are shown in Table 1. We did not identify any relevant observational studies.

Risk of bias and results of RCTs

Risk of bias assessment was based on the main RCT results paper of the included trials (Table 2). The method of randomisation was explicit (low risk) in all four RCTs. Allocation concealment was adequate (low risk) in all trials. None of the trials were described as double blinded (participant or therapist) and rated as high risk; however in one study participants were blinded to treatment as sham treatment used in one arm [20] and in another study researchers doing assessments were

Table 2
Cochrane collaboration risk of bias assessment (Higgins and Green, Cochrane handbook for systematic reviews of interventions, 2011).

Study ID	Risk of Bias Assessment (RoBA)*								
	1	2	3	4	5	6	7	8	Funder
BeST Trial, United Kingdom (Lamb et al., HTA, 2010;14; Underwood et al., Arthritis Care Res, 2011;63:1271)	L	L	H	L	L	L	L	L	NIHR HTA Programme
Cherkin Trial, USA (Cherkin et al., Arch Intern Med, 2009;169:858; Sherman et al., BMC Musculoskelet Disord, 2009;10:114)	L	L	H	L	L	L	L	L	NIH Cooperative agreement with National Centre for Complementary and Alternative Medicine
UK BEAM Trial, United Kingdom (UK BEAM, BMJ, 2004;329:1377; Underwood et al., Rheumatology, 2007;46:1297)	L	L	H	L	L	L	L	L	Research Costs: Medical Research Council Treatment Costs: NHS, Research and Development (R&D)
Witt Trial, Germany (Witt et al., Clin J Pain, 2011;27:550)	L	L	H	U	U	L	U	H	A group of social health fund providers

* RoBA, (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) incomplete outcome data, (5) selective reporting, (6) similarity of groups at baseline, (7) sample size calculation, (8) intention to treat analysis. L—low risk of bias, H—high risk of bias, U—unclear.

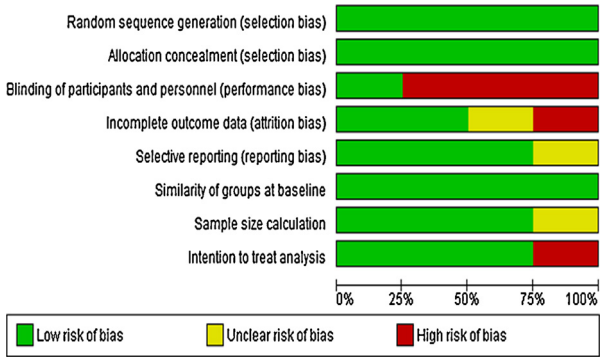

Fig. 1. Cochrane collaboration risk of bias assessment tool.

masked [21]. Three trials had no evidence of selective outcome reporting and dropout rate analyses were adequately prevented and were judged as being low risk [3,25,26]. The same three trials included an intention to treat (ITT) analysis and had adequate sample size based on power calculations for the main comparison. Only one trial did not carry out an ITT analysis and provided unclear evidence of selective outcome reporting [24] (Fig. 1). BEST trial was funded by NIHR HTA programme, Cherkin trial by NIH cooperative agreement with National centre for complementary and Alternative Medicine, UK BEAM Trial by Medical research council and NHS, Research and development and Witt trail by a group of social health fund providers.

Methodological quality for subgroups

The methodological quality of the moderator analyses varied; BeST provided confirmatory evidence for fulfilling all five criteria for subgroup studies for two potential moderators [23]. Cherkin and UK BEAM provided exploratory evidence i.e. they only met criteria three, four and five for subgroup studies [7,22] and Witt provided insufficient data to judge quality of subgroup analyses [24] (Table 3).

Moderator variables identified

Potential moderators with strong evidence ($p < 0.05$) in one or more studies include age (younger participants may gain more benefit), employment status and type (those employed or in sedentary occupations may gain greater benefit), back pain status (those who are worse may gain greater benefit), narcotic medication use (users may benefit less), treatment expectations (those with a greater positive expectation gained more benefit) and education (those with greater than 10 years of schooling gained a greater benefit). Potential moderators with weaker evidence ($0.05 < p \leq 0.20$) include gender (female participants may gain greater benefit), psychological distress (those with anxiety and depressive symptoms may benefit more), pain/disability (those with greater pain/disability at baseline may benefit more) and quality of life (those with a better quality of life may benefit more) (Table S1).

Interaction with age was found in the BeST, Cherkin and Witt trials [5,20,24]. Specifically a cognitive behavioural approach was more beneficial in younger participants than older participants on the RMDQ score. The evidence for this was strong, with a treatment difference of −1.58 ($p = 0.035$; 95% CI −3.05 to −0.12). Witt [24] also found a statistically significant additional benefit from acupuncture treatment in younger participant ($p < 0.001$).

A cognitive behavioural approach produced a comparatively greater improvement in females compared to males in the BeST trial [5]. For the RMDQ score the treatment difference between male and female was −1.27 ($p = 0.102$; 95% CI −2.79 to 0.25), providing some weak evidence.

Table 3
Quality assessment of the treatment moderators (Pincus et al., BMC Med Res Methodol, 2011;11:14).

Study ID	Was the subgroup analysis specified a-priori	Was the selection of subgroup factors for analysis theory/evidence driven	Were subgroup factors measured prior to randomisation	Was measurement of subgroup factors measured by adequate (reliable and valid) measurements, appropriate for the target population	Does the analysis contain an explicit test of the interaction between moderator and treatment	Strength of evidence
BeST (Lamb et al., HTA, 2010;14; Underwood et al., Arthritis Care Res, 2011;63:1271)	Yes	Yes	Yes	Yes	Yes	Confirmatory evidence (for two variables only)
Cherkin (Cherkin et al., Arch Intern Med, 2009;169:858; Sherman et al., BMC Musculoskelet Disord, 2009;10:114)	No	No	Yes	Yes	Yes	Exploratory evidence
UK BEAM (UK BEAM, BMJ, 2004;329:1377; Underwood et al., Rheumatology, 2007;46:1297)	No	No	Yes	Yes	Yes	Exploratory evidence
Witt (Witt et al., Clin J Pain, 2011;27:550)	No	No	Unclear	No	Unclear	Insufficient evidence

Confirmatory evidence: The study fulfilling all five criteria for moderator studies; *Exploratory evidence*: The study meeting only three, four and five criteria for moderator studies; *Insufficient evidence*: The study not meeting criteria of explicit test of interaction between moderator and treatment and inadequate measurement of subgroup factors.

Being employed and the type of employment had some moderating effect. In the BeST trial [5] a cognitive behavioural approach produced additional benefit in employed participants when compared to those who were not employed. Greater education also had a beneficial effect on treatment outcome. In Witt [24], participants who have had more than 10 years of schooling gained a greater benefit from acupuncture ($p = 0.01$). BeST [5] found a treatment difference of 1.29 ($p = 0.098$; 95% CI −0.24 to 2.82) on the RMDQ score for participants leaving education after the age of 16, this provides some weak evidence.

Manipulation treatment provided greater benefit at three months ($p = 0.176$) and 12 months ($p = 0.143$) for the RMDQ score amount those with greater pain/disability at baseline [3]. The evidence for this is weak. Worse initial back pain status also produced greater gain from acupuncture in the Cherkin and Witt trials [20,24] when compared to those with a better back pain status at baseline (p-values ranged from <0.001 to 0.16). There was a weak interactions for how troublesome or bothersome back pain was perceived as, where greater benefit from treatment was in those with a more troublesome/bothersome condition.

Having better expectations about the treatment was found to be a moderating factor with p-values ranging between $p = 0.03$ and $p = 0.192$ demonstrating a spectrum of strong to weak evidence for the interactions.

Baseline anxiety and depression had a weak moderating effect. Greater baseline anxiety resulted in more benefit from treatment in terms of the RMDQ score, the treatment difference was −1.12 ($p = 0.195$; 95% CI −2.83 to 0.58). Those with higher levels of depression gained more benefit from the treatment than those who were less depressed for outcome of RMDQ and MVK disability score. The treatment difference was found to be −2.07 ($p = 0.135$; 95% CI −4.79 to 0.65) and −14.58 ($p = 0.051$; 95% CI −29.19 to 0.03) for the RMDQ and MVK disability score, respectively.

Discussion

The aim of this review was to identify variables from current evidence that are potential moderators of treatment effect; variables that have shown to have a possible beneficial moderating effect (at a ≤0.20 level of significance) on treatments for LBP. We were only able to include data from four trials; two of these had considered subgroup analysis at the design stage and only one of these (Witt) was powered to show such an effect. Despite this, the Witt trial ranked poorly on the quality assessment tools therefore caution should be taken when interpreting or applying the findings. Across these four trials a large number of moderator analyses were performed; not all of which were presented in the published papers. We have only presented a small number of interactions, some statistically significant and others approaching statistical significance. The likelihood of a statistically significant finding increases when a large number of tests are carried out. Thus caution is needed in interpreting the clinical importance of any moderation that we have identified; particularly as none of the trials were testing similar interventions.

As moderator variables have not been looked for in this way before it is hard to compare our findings to that of others. Our findings do seem to concur with the findings of some of the earlier work in this area particularly in the moderating effects of age and employment, moderate disability and fear avoidance belief on treatment [27,28]. We identified one further trial which, prima facia, reported that intervention effects differed with gender; but the analysis did not include a formal test for interactions so we excluded it [29]. Our findings about treatment expectations also concur with earlier work [30–32]. A systematic review of factors that influenced outcome from self-management programmes for chronic musculoskeletal pain found that self-efficacy, depression, pain catastrophising and physical activity were important [33].

In an individual patient data meta-analysis of approximately 9990 participants with LBP, headache, neck pain or osteoarthritis it was found that gender, the living situation of the patients, earlier positive acupuncture treatments and a failure of other therapies were potential effect moderators [24]. We identified one additional publication using the UK BEAM data that suggested that a moderator analysis had been done [34]. It did not, however, report interaction between treatment group and work status or educational group or Townsend score separately, rather it had combined them in a single analysis; therefore we excluded it. The STarT Back trial compared the overall effectiveness of an approach using prognostic stratification and matched interventions for low, medium and high risk subgroups, with usual best care for back pain after our searches had been completed. However, interactions for individual variables were not reported and it would therefore not have been eligible for inclusion in our review [35].

Strengths and limitations

The main strength of this study is that, to our knowledge, this is the first review to have looked at all papers related to large trials that may plausibly show an interaction rather than searching for subgroup effects. We only included secondary analysis data RCTs that were large enough to detect a moderate standardised effect size for moderation of 0.5. This ensured the credibility of the moderators identified from these studies for the purposes of our review. We carried out a comprehensive systematic search separately for RCTs and observational studies to maximise identification of all published studies related to low back pain. However we did not find any observational studies for inclusion. We used strict inclusion and exclusion criteria to ensure high quality of the included studies. None of the observational studies met our inclusion criteria. We used two quality assessment tools - 'Cochrane risk of bias assessment tool' and 'methodological tool for treatment moderators', [18] to assess quality and level of evidence of the included studies.

We could not test for publication bias because we have included only four studies in our review. This meant that the power of the test was too low to distinguish chance from real asymmetry [17] and statistical heterogeneity [36]. We presented secondary analysis and results of subgroup-specific analysis as reported in the published papers; hence the findings should be interpreted with caution, as there is a possibility of bias associated to many factors such as inappropriate statistical methods and insufficient a priori specification of variables [37].

We included RCTs in which participants or therapist cannot be blinded/masked to treatment arm because of the nature of the studies. However, these were all well-conducted RCTs with adequate concealment of allocation and adequate generation of the allocation sequence. In Witt et al., the only evidence of a moderator effect came from a statement reported in the paper [24]. We contacted the authors to clarify this point but they did not provide the actual data only the *p*-values.

The findings from our review provide some very weak empirical evidence that certain groups of patients might derive a greater benefit from therapist delivered interventions. The evidence is not strong enough to make clinical recommendations. The STarT Back trial found a stratified management approach with prognostic screening and targeted treatments to be clinically and cost-effective [35] reinforcing the need to develop an understanding of the characteristics of patients who benefit the most from a given treatment.

There are still arguments, for and against, subgrouping [10]. It is unlikely that any single trial will be sufficiently resourced to be statistically powered to do subgroup analyses of all possible moderating variables; and no guarantee that trials included in any individual patient data meta-analysis will be sufficiently homogenous to allow sub-group identification. It may be that seeking to identify subgroups in clinical trial data will fail to produce a useful clinical classification. In which case the back pain research community should consider developing clinically defined subgroups in which different interventions can be tested. Our challenge now is to explore the variables we have identified in individual patient data (IPD) meta-analyses. However, we recognise as part of this challenge that we have, almost certainly, not identified all of the potential moderators.

Conclusion

This study provides some insight into the potential moderators with strong ($p < 0.05$) and weak ($0.05 < p \leq 0.20$) evidence. There are however insufficient robust data on moderators to be useful in clinical practice. This review has identified some important potential moderators of treatment effect worthy of testing in future confirmatory analyses although some caution is needed in interpreting the findings.

Funding

This paper summarises independent research funded by the National Institute for Health Research (NIHR) under

its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0608-10076). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflict of interest statement

None.

Acknowledgments

This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0608-10076). The views expressed in this paper are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

This project benefitted from facilities funded through Birmingham Science City Translational Medicine Clinical Research and Infrastructure Trials Platform, with support from Advantage West Midlands (AWM) and the Wolfson Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.physio.2015.01.006.

References

[1] Maniadakis N, Gray A. The economic burden of back pain in the UK. Pain 2000;84(1):95–103.
[2] NICE. Early management of persistent non-specific low back pain. London: NICE; 2011.
[3] UK BEAM. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. BMJ 2004;329(7479):1377.
[4] NICE. Briefing paper for methods review workshop on identifying subgroups and exploring heterogeneity. London: NICE; 2007.
[5] Lamb SE, Lall R, Hansen Z, Castelnuovo E, Withers EJ, Nichols V, *et al.* A multicentred randomised controlled trial of a primary care-based cognitive behavioural programme for low back pain. The back skills training (BeST) trial. HTA 2010;14(41):1–281.
[6] Tilbrook HE, Cox H, Hewitt CE, Kang'ombe AR, Chuang LH, Jayakody S, *et al.* Yoga for chronic low back pain: a randomized trial. Ann Intern Med 2011;155(9):569–78.
[7] Underwood MR, Morton V, Farrin A, UK Beam Trial Team. Do baseline characteristics predict response to treatment for low back pain? Secondary analysis of the UK BEAM dataset. Rheumatology 2007;46(8):1297–302.
[8] Henschke N, Maher CG, Refshauge KM, Das A, McAuley JH. Low back pain research priorities: a survey of primary care practitioners. BMC Fam Pract 2007;8:40.
[9] Savigny P, Watson P, Underwood M. Early management of persistent non-specific low back pain: summary of NICE guidance. BMJ 2009;338:b1805.
[10] Foster NE, Hill JC, Hay EM. Subgrouping patients with low back pain in primary care: are we getting any better at it? Man Ther 2011;16(1):3–8.
[11] Kent PM, Keating JL. Can we predict poor recovery from recent-onset nonspecific low back pain? A systematic review. Man Ther 2008;13(1):12–28.
[12] Pengel LH, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. BMJ 2003;327(7410):323.
[13] Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. Spine 2002;27(5):E109–20.
[14] Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002;59(10):877–83.
[15] Turner JA, Holtzman S, Mancl L. Mediators, moderators, and predictors of therapeutic change in cognitive-behavioral therapy for chronic pain. Pain 2007;127(3):276–86.
[16] Kamper SJ, Maher CG, Hancock MJ, Koes BW, Croft PR, Hay E. Treatment-based subgroups of low back pain: a guide to appraisal of research studies and a summary of current evidence. Best Pract Res Clin Rheumatol 2010;24(2):181–91.
[17] Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
[18] Pincus T, Miles C, Froud R, Underwood M, Carnes D, Taylor SJ. Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. BMC Med Res Methodol 2011;11:14.
[19] Lachenbruch PA. A note on sample size computation for testing interactions. Stat Med 1988;7(4):467–9.
[20] Cherkin DC, Sherman KJ, Avins AL, Erro JH, Ichikawa L, Barlow WE, *et al.* A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. Arch Intern Med 2009;169(9):858–66.
[21] Lamb SE, Hansen Z, Lall R, Castelnuovo E, Withers EJ, Nichols V, *et al.*, Back Skills Training Trial. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. Lancet 2010;375(9718):916–23.
[22] Sherman KJ, Cherkin DC, Ichikawa L, Avins AL, Barlow WE, Khalsa PS, *et al.* Characteristics of patients with chronic back pain who benefit from acupuncture. BMC Musculoskelet Disord 2009;10:114.
[23] Underwood M, Mistry D, Lall R, Lamb S. Predicting response to a cognitive-behavioral approach to treating low back pain: secondary analysis of the BeST data set. Arthritis Care Res 2011;63(9):1271–9.
[24] Witt CM, Schutzler L, Ludtke R, Wegscheider K, Willich SN. Patient characteristics and variation in treatment outcomes: which patients benefit most from acupuncture for chronic pain? Clin J Pain 2011;27(6):550–5.
[25] Cherkin DC, Sherman KJ, Hogeboom CJ, Erro JH, Barlow WE, Deyo RA, *et al.* Efficacy of acupuncture for chronic low back pain: protocol for a randomized controlled trial. Trials 2008;9:10.
[26] Lamb SE, Lall R, Hansen Z, Withers EJ, Griffiths FE, Szczepura A, *et al.* Back skills training trial. Design considerations in a clinical trial of a cognitive behavioural intervention for the management of low back pain in primary care: Back Skills Training Trial. BMC Musculoskelet Disord 2007;8:14.
[27] Staal JB, Hlobil H, Koke AJ, Twisk JW, Smid T, van MW. Graded activity for workers with low back pain: who benefits most and how does it work? Arthritis Rheum 2008;59(5):642–9.
[28] Steenstra IA, Knol DL, Bongers PM, Anema JR, van MW, de Vet HC. What works best for whom? An exploratory, subgroup analysis in a randomized, controlled trial on the effectiveness of a workplace intervention in low back pain patients on return to work. Spine 2009;34(12):1243–9.
[29] Becker A, Leonhardt C, Kochen MM, Keller S, Wegscheider K, Baum E, *et al.* Effects of two guideline implementation strategies on patient outcomes in primary care: a cluster randomized controlled trial. Spine 2008;33(5):473–80.

[30] Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA. Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. Spine 2001;26(13): 1418–24.
[31] Myers SS, Phillips RS, Davis RB, Cherkin DC, Legedza A, Kaptchuk TJ, *et al.* Patient expectations as predictors of outcome in patients with acute low back pain. J Gen Intern Med 2008;23(2):148–53.
[32] Thomas KJ, MacPherson H, Thorpe L, Brazier J, Fitter M, Campbell MJ, *et al.* Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. BMJ 2006;333(7569):623.
[33] Miles CL, Pincus T, Carnes D, Homer KE, Taylor SJ, Bremner SA, *et al.* Can we identify how programmes aimed at promoting self-management in musculoskeletal pain work and who benefits? A systematic review of sub-group analysis within RCTs. Eur J Pain 2011;15(8), 775-782.
[34] Moffett JA, Underwood MR, Gardiner ED. Socioeconomic status predicts functional disability in patients participating in a back pain trial. Disabil Rehabil 2009;31(10):783–90.
[35] Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, *et al.* Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet 2011;378(9802):1560–71.
[36] Centre for Dissemination. Guidance for undertaking systematic reviews in health care. Centre for Dissemination; 2009.
[37] Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey SG. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. HTA 2001;5(33):1–56.