

Original article

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

Majid Artus¹, Danielle A. van der Windt¹, Kelvin P. Jordan¹ and Elaine M. Hay¹

Abstract

Objectives. To assess overall responses to treatments among non-specific low back pain (NSLBP) patients in clinical trials to examine the pattern following a wide range of treatments.

Methods. We conducted a systematic review of published trials on NSLBP and meta-analysis of within-group responses to treatments calculated as the standardized mean difference (SMD). We included randomized controlled trials that investigated the effectiveness of primary care treatments in NSLBP patients aged ≥ 18 years. Outcome measures included the visual analogue scale for pain severity, Roland Morris Disability Questionnaire and Oswestry Disability Index for physical functioning.

Results. One hundred and eighteen trials investigating a wide range of primary care treatment for NSLBP were included. Plots of response to treatments showed that there was a similar pattern of initial improvement at 6 weeks followed by smaller improvement for both pain and functional disability at long-term follow-up. This was also shown by the pooled SMD for pain which was 0.86 (95% CI 0.65, 1.07) at 6 weeks, 1.07 (95% CI 0.87, 1.27) at 13 weeks, 1.03 (95% CI 0.82, 1.25) at 27 weeks and 0.88 (95% CI 0.60, 1.1) at 52 weeks. There was a wide heterogeneity in the size of improvement. This heterogeneity, however, was not explained by differences in the type of treatment classified as active, placebo, usual care or waiting list controls or as pharmacological or non-pharmacological treatment.

Conclusions. NSLBP symptoms seem to improve in a similar pattern in clinical trials following a wide variety of active as well as inactive treatments. It is important to explore factors other than the treatment, that might influence symptom improvement.

Key words: Systematic review, Low back pain, Non-specific effects, Primary care.

Introduction

Low back pain is a common and costly condition. Estimates of the lifetime prevalence of non-specific low back pain (NSLBP) range from 49 to 70% [1], indicating that most people will experience one or more episodes of back pain during their lifetime. Although many episodes of acute low back pain resolve rapidly [2], ~30% result in persistent and relapsing, disabling symptoms [3, 4]

leading to large direct and indirect health care costs [5, 6]. In the UK, most patients with back pain are managed within primary health care with 7–9% of the UK population consulting a general practitioner with an episode of back pain during the course of a year [7].

Interest in evaluating the effectiveness of primary care treatments for back pain has grown substantially over the past two decades as reflected in the increasing number of publications of randomized trials and systematic reviews [8–10]. However, most trials and reviews are unable to show a clear superiority or benefit of any particular treatment, with different treatments apparently leading to a similar and significant improvement in symptoms and with differences between treatment effects that are often modest [11–17]. Changes in outcome scores within

¹Arthritis Research UK National Primary Care Centre, Primary Care Sciences, Keele University, Keele, Staffordshire, UK.

Submitted 5 February 2010; revised version accepted 25 June 2010.

Correspondence to: Majid Artus, Arthritis Research UK National Primary Care Centre, Primary Care Sciences, Keele University, Keele, Staffordshire ST5 5BG, UK. E-mail: m.artus@cphc.keele.ac.uk

treatment groups in these trials, which represent overall symptom improvement, suggest a pattern of improvement occurring early after onset of treatment and possibly unrelated to the type of treatment used [18, 19].

However, we are not aware of any conclusive evidence for these observations based on a large number of trials across varied types of treatments for low back pain. The objective of our review was to assess overall responses to treatments among NSLBP patients in clinical trials to examine the pattern following a wide range of treatments. We included a wide variety of treatments ranging from simple advice to intensive multidisciplinary rehabilitation pain management programmes including examples such as medications, acupuncture, chiropractic and transcutaneous electrical nerve stimulation (TENS). The aim of our review was not to study the effectiveness of these treatments, but to investigate the pattern of symptom improvement within intervention groups following their use. This was performed by conducting a systematic review of published trials on NSLBP and meta-analysis of overall reduction in pain and functional disability.

Methods

Inclusion criteria and data source

Randomized controlled trials that investigated the effectiveness of primary care treatments in patients with NSLBP aged ≥ 18 years were included. NSLBP was defined as pain in the area below the lower ribs and above the gluteal folds with no known underlying pathology [20]. We excluded trials conducted in patients with low back pain of identifiable cause (e.g. disc herniation or inflammatory arthritis such as AS), post-operative or post-traumatic back pain, or back pain during pregnancy or labour. Also excluded were trials conducted among healthy volunteers and those published in languages other than English.

As the emphasis was on the type of treatment rather than the setting of the trial, trials evaluating primary care treatments in settings other than primary care were also included. Primary care treatments were defined as those that would usually be within the expertise of primary health care practitioners and provided within the usual facilities of primary health care or their equivalents such as physiotherapy departments, rehabilitation units and occupational health care departments. Examples of treatments include Yoga, non-steroidal anti-inflammatory tablets, exercises, chiropractic adjustment, physical therapy, TENS, acupuncture and osteopathic manipulation (supplementary appendix A, available as supplementary data at *Rheumatology Online*). Examples of primary care practitioners include general practitioners, family doctors, physiotherapists, occupational therapists, acupuncturists, osteopaths and psychologists.

We chose outcome measures commonly used in the majority of NSLBP trials, namely the visual analogue scale (VAS, 10 cm or 100 mm) or the numerical rating scale (NRS, 0–10) for pain severity [21], the Roland Morris Disability questionnaire (RMDQ, 24 point) [22] or

its modified versions or Oswestry Disability Index (ODI, 0–100) [23] for physical functioning.

To satisfy the aims of this review, we needed to identify trials that investigated a wide range of primary care treatments for NSLBP. We searched the Cochrane Register of Controlled Trials' (CENTRAL) database first quarter issue of 2007 [24] using the term low back pain. The quality of included trials was assessed [25] and the findings and their association with response to treatments will be presented elsewhere.

Data analysis

To enable comparison between trials, five commonly used follow-up time points were selected (6, 13, 27, 52 and 104 weeks) and data were analysed if they were within 3 weeks of these time points. When baseline and follow-up outcome scores had not been reported using mean and s.d., for example, medians and interquartile ranges, we converted them into mean and s.d. using the method of Hozo *et al.* [26].

Examining responses to treatments for their pattern was made in three steps including a descriptive assessment of trend in responses, assessment of variation in size of response and summarizing the overall pattern of responses as follows.

Describing the general pattern of responses to treatment

Exploration of the general pattern of response was first done through visual assessment of outcome scores on pain and functional disability. These were plotted as graphic lines representing baseline and follow-up mean scores at each time point for each arm of each trial. These response lines were plotted using Excel 2000 (Microsoft, Redmond, WA, USA).

Assessing variation in size of response to treatment

To examine variation in the size of response to treatment (heterogeneity) across trial arms, changes in outcome scores were analysed by calculating standardized mean difference (SMD) [27]. The SMD was calculated for each trial arm by subtracting the follow-up mean score on the outcome measure from the baseline mean score and dividing by its s.d. at baseline. This was done separately for each of the three outcome measures (RMDQ, ODI and pain rating). The 95% CIs for the SMDs for each measure for each trial arm were also calculated [28].

The SMD is a method of standardizing the measurement of change over time so that studies using slightly different scales, but measuring the same underlying construct, can be combined and more easily compared. In our case, this meant that by using the SMD, studies which used the modified versions of the RMDQ could be included in the same analysis as those which used the original version. Similarly, studies which used a VAS for pain could be combined into the same analysis as those using an NRS.

Although the RMDQ and ODI measure similar constructs, studies using the ODI were not included in the same analysis as the RMDQ as their differing content and any difference in their properties (such as different

levels of responsiveness) would add a potential extra element of heterogeneity. However, by using SMDs, comparisons in pattern of responses to treatment between these measures can be more easily made.

The SMDs, stratified by measurement (RMDQ, ODI and pain rating), were plotted in forest plots representing with their 95% CIs. Heterogeneity in outcome between trial arms for each measure was investigated by computing the I^2 -statistic [29]. The I^2 -statistic quantifies the effect of heterogeneity describing the percentage of total variation in outcome between trial arms that cannot be explained by chance (0% = no heterogeneity, 100% = very significant heterogeneity).

Summarizing the overall pattern of responses to treatments

In order to summarize the overall responses to treatment across trial arms, we calculated a combined pooled estimate of SMD for each follow-up time point for each measure using a random effects model weighted by inverse variance [30].

The pooled analyses were carried out with one arm randomly selected from each trial. This was to overcome any potential similarity or clustering of responses between the arms of each trial. Participants from the same trial are more likely to have similar characteristics, and respond in the same way to treatment as participants from other trials, and we wanted to examine the pattern of responses without the potential influence of this clustering. A sensitivity analysis was carried out to compare these results with the results of an analysis based on all trial arms.

Subgroup analyses

We next explored the evidence for the hypothesis that responses follow a common pattern regardless of the type of treatment. We therefore repeated the analysis for stratified subgroups of trials. Trial arms were stratified according to the type of treatment using two classifications: (i) pharmacological, non-pharmacological, other and mixed treatments and (ii) index treatment, active comparator treatment, usual care, waiting list or placebo treatment.

We also stratified trial arms into those that were conducted among the general population and those that were conducted in other subsettings (general practice, physiotherapy departments and occupational therapy units). Given the likely difference in baseline severity, this would enable us to explore any possible association between baseline pain and disability, and rate of improvement. Analysis was performed using STATA/IC 10.0 software (Stata Corp., College Station, TX, USA).

Results

Trials' description

Our search of CENTRAL yielded a total of 772 trial citations. One hundred and twenty-six trials satisfied our inclusion criteria (supplementary appendix A, available as supplementary data at *Rheumatology Online*).

Eight published papers were each a second report of the same trial and therefore both were used to extract data related to that trial. Data, therefore, were available from 118 trials (Fig. 1).

Trial sample size ranged from 20 to 719 and duration of follow-up from 5 days to 3 years. Participants' age ranged from 27 to 79 years (median of 42 years) and the range of percentage of females among participants was 0–100% (median of 55%). Primary health care subsettings included general practice subsetting (29 trials), occupational health care departments (20 trials) and physiotherapy departments (10 trials). Fourteen trials were conducted among the general population, 31 in mixed settings and in 14 trials the setting could not be clearly identified. Fifty-seven (48%) trials were conducted among patients with chronic low back pain, 27 (23%) trials among acute low back pain patients, 23 (20%) trials among subacute low back pain patients, 8 (7%) trials among patients with acute and subacute or subacute and chronic low back pain, and in 3 trials the type of pain was not clear.

The majority of trials included in this review used another active treatment as the comparator (79 trials, 67%) and in 27 (34%) of these trials more than one active comparator treatments were used. Placebo or sham treatment arms were used in 36 (31%) trials and no-treatment arms (waiting list control) in 11 (9%) trials. Ninety-one (77%) trials were conducted to assess non-pharmacological treatments [27 (30%) of which assessed treatments with a psychological component], 20 (17%) trials investigated pharmacological treatments, 5 (4%) trials mixed treatments, whereas it was not possible to classify two trials according to the type of treatment used according to this classification.

The pattern of responses to treatments

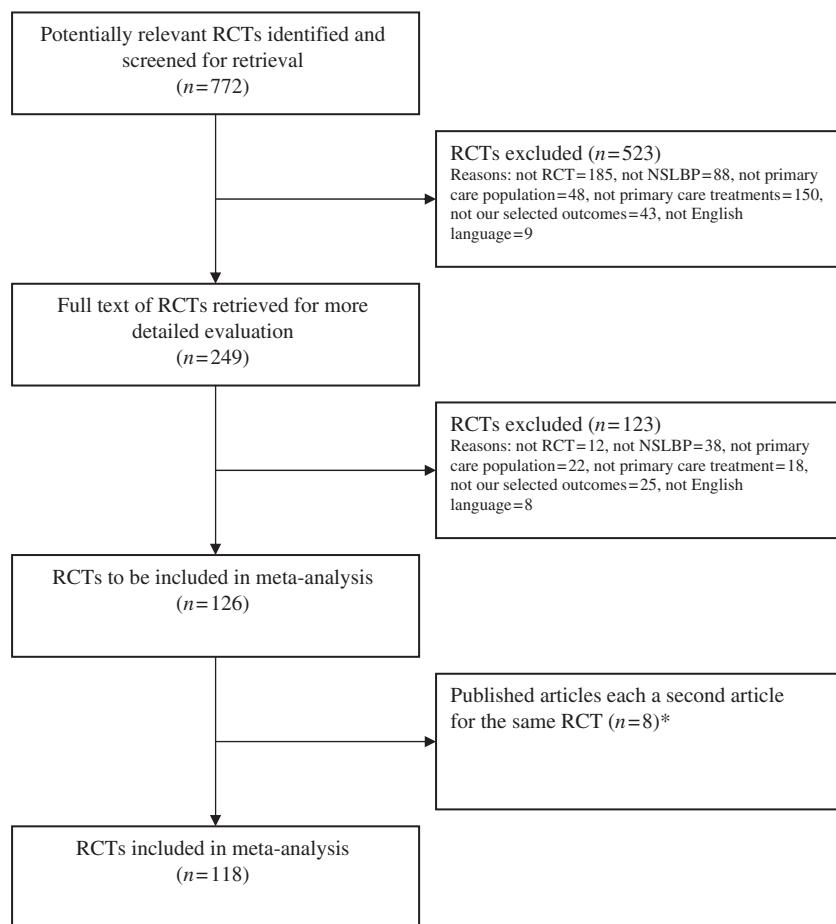
Data for pain severity outcome were available from 104 treatment arms in 45 trials, for RMDQ outcome from 82 treatment arms in 35 trials and for ODI outcome from 61 treatment arms in 26 trials. A graphical representation of overall responses for each treatment in each trial arm up to 12-month follow-up for pain severity is presented in Fig. 2. Responses for RMDQ and ODI showed similar patterns (supplementary appendix B, available as supplementary data at *Rheumatology Online*).

Response lines for all three outcome measures followed a pattern of common trend of improvement in symptoms represented by a rapid early reduction in mean outcome scores within the first 6 weeks followed by a slower reduction thereafter proceeding to a plateau at 6 months. This common trend in responses remained when responses from only one arm were randomly selected from each trial. This pattern appeared to be similar regardless of the type of treatment received.

Variation in responses to treatments

Figure 3 shows a forest plot representing SMDs and 95% CIs for pain at the 27-week follow-up point for one randomly selected arm from each trial. It shows wide heterogeneity in effect size between trial arms; however,

Fig. 1 Flow diagram for inclusion of trials for the review. *When two articles were published for the same trial, information was extracted from both and considered as one trial. RCTs: randomized clinical trials.



the majority of trial arms showed improvement in symptoms. A similar variation in size of response was found for both functional disability outcomes and at each follow-up point. The I^2 -statistic confirmed the wide heterogeneity in SMDs between trial arms (generally >80%).

Summary of responses to treatments

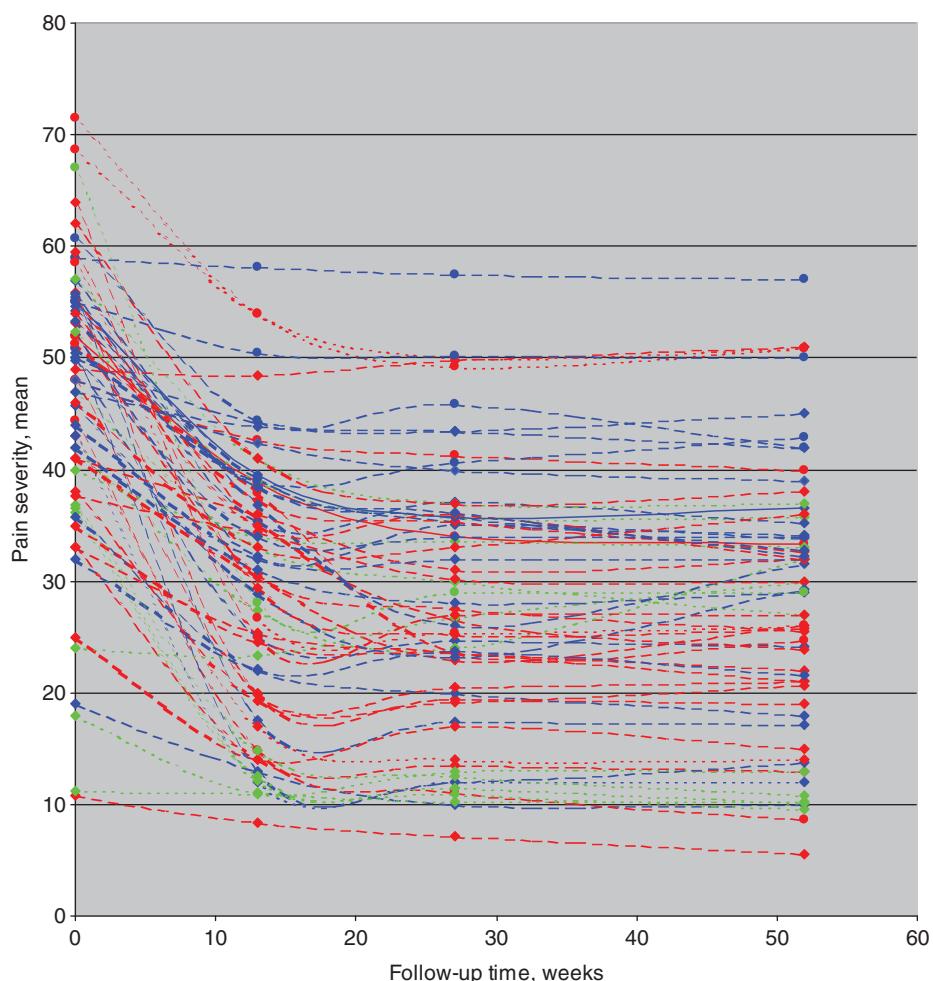
A random effects model was used to compute pooled SMDs from one randomly selected arm from each trial to explore the overall trend in changes in pain and functional disability over time (Table 1). The pattern of responses, which is illustrated graphically in Fig. 2, was confirmed by common large initial SMDs at 6-week follow-up after which only smaller further improvements were seen. For pain, pooled SMD was 0.86 (95% CI 0.65, 1.07) at 6 weeks, 1.07 (95% CI 0.87, 1.27) at 13 weeks, 1.03 (95% CI 0.82, 1.25) at 27 weeks and 0.88 (95% CI 0.60, 1.1) at 52 weeks. Similar findings were produced when all arms of included trials were included (Table 2). Overall responses were large, as SMDs >0.8 are considered as large; 0.5–0.8 moderate and <0.5 small [31].

Results of subgroup analyses

Variation in treatment responses did not appear to be explained by different types of treatment. Table 3 shows that there were similar overall responses in treatment groups receiving different types of treatment (index, active comparator, usual care, placebo or waiting list control) and in groups being treated with pharmacological or non-pharmacological treatments.

Some of the trials included in our review were conducted among the general population. The mean severity of pain was lower in these trials than that in trials conducted in other subset settings [mean baseline scores 39.7 (s.d. 18.2) compared with 54.7 (s.d. 22.1), 50.3 (s.d. 23.4) and 44.7 (s.d. 20.5) for trials in general practice, occupational health care and physiotherapy departments, respectively]. Pooled SMD for pain at 27 weeks for general population trials was 0.89 (95% CI 0.67, 1.11) and for the other three subset settings 1.25 (95% CI 0.94, 1.56), 1.03 (95% CI 0.74, 1.32) and 1.34 (95% CI 1.11, 1.57). Although there was a wide heterogeneity in SMDs between trials, we found that the percentage change in symptom improvement between baseline and final

Fig. 2 Overall responses (VAS for pain) up to 52-week follow-up in each treatment arm of included trials. Each line represents a response line within each trial arm. Red: index treatment arm; Blue: active treatment arm; Green: usual care/waiting list/placebo arms. _____: pharmacological treatment; - - - : non-pharmacological treatment;: mixed/other.



follow-up remained the same regardless of pain severity at baseline. Baseline mean pain scores for all included treatment arms was 47.0 (s.d. 11.9). Trial arms with baseline mean pain scores below and above this average showed a mean relative rate of improvement at 27-week follow-up of 47 and 48%, respectively.

Discussion

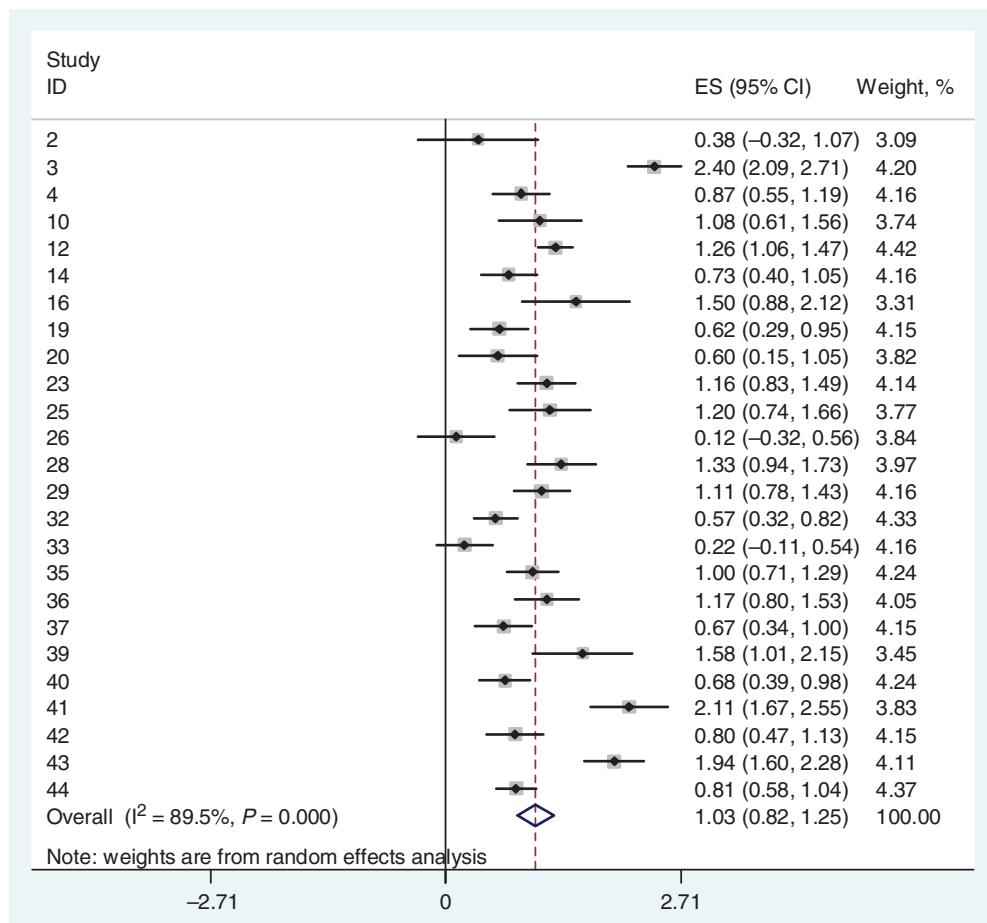
The main aim of this review was to summarize evidence on responses to treatment among low back pain patients in clinical trials in order to test the hypothesis that these responses follow a similar pattern regardless of the treatment used.

We examined responses from trials in which patients varied in the duration and severity of their pain and were conducted over a period spanning 15 years and in more than 18 countries. A wide variety of active treatments were included ranging from tablet medications,

psychological treatment and simple advice to hands on manual therapies and extensive multidisciplinary pain management programmes.

In this review, we found evidence that these responses seem to follow a common trend of early rapid improvement in symptoms that slows down and reaches a plateau 6 months after the start of treatment, although the size of response varied widely. We found a similar pattern of improvement in symptoms following any treatment, regardless of whether it was index, active comparator, usual care or placebo treatment. To understand the meaning of these findings and draw any useful conclusions, we need to explore several explanations.

Outcome in trials will be influenced by the intervention itself (specific effect), non-specific factors, random variation and errors in the trial design or analysis. Non-specific factors may include natural history, regression to the mean and non-specific effects of treatment. How these may affect response is now discussed.

FIG. 3 SMDs for pain for one arm randomly selected from each trial at 27-week follow-up. ES: effect size.**TABLE 1** SMDs (within-group responses to treatment) for pain and disability (RMDQ and ODQ) for one arm randomly selected from each trial for the main follow-up time points

Outcome	Follow-up time point, weeks	Number of trials	Total number of arms	Number of randomly selected arms	SMD		
					Pooled ^a (95% CI)	Range	I^2 , %
Pain	6	13	27	13	0.86 (0.65, 1.07)	0.21, 2.00	82
	13	29	58	29	1.07 (0.87, 1.27)	0.09, 4.25	91
	27	25	50	25	1.03 (0.82, 1.25)	0.12, 2.11	90
	52	23	48	23	0.88 (0.60, 1.11)	-0.30, 2.44	92
	104	5	11	5	0.59 (0.45, 0.74)	-0.39, 0.98	91
RMDQ	1	8	16	8	0.97 (0.75, 1.19)	0.51, 1.47	69
	6	19	39	19	0.97 (0.66, 1.28)	-0.17, 2.22	95
	13	21	42	21	0.93 (0.67, 1.20)	-0.12, 2.39	94
	27	12	24	12	0.91 (0.59, 1.24)	0.05, 1.98	93
	52	11	22	11	1.01 (0.68, 1.34)	0.13, 2.04	92
ODQ	1	5	12	5	0.92 (0.59, 1.24)	0.63, 1.35	58
	6	14	30	14	0.98 (0.62, 1.33)	-0.09, 3.40	90
	13	12	24	12	0.92 (0.70, 1.14)	0.39, 3.50	81
	27	10	20	10	1.08 (0.80, 1.36)	0.45, 1.66	83
	52	12	24	12	1.14 (0.88, 1.39)	0.61, 2.63	84
	104	4	8	4	1.05 (0.57, 1.54)	0.54, 1.80	90

^aPooled SMDs (using random effects model) of one arm randomly selected from each trial.

TABLE 2 SMDs (within-group responses to treatment) for pain and disability (RMDQ and ODQ) for all trial arms for the main follow-up time points

Outcome	Follow-up time point, weeks	Number of trials	Total number of arms	SMD		I^2 , %
				Pooled ^a (95% CI)	Range	
Pain	6	13	27	0.83 (0.70, 0.96)	0.08, 2.13	82
	13	29	58	0.99 (0.86, 1.11)	0.09, 4.25	91
	27	25	50	1.07 (0.93, 1.20)	0.05, 2.45	90
	52	23	48	0.91 (0.76, 1.05)	-0.09, 2.45	89
	104	5	11	0.77 (0.50, 1.03)	-0.39, 1.36	87
RMDQ	1	8	16	0.80 (0.63, 0.98)	0.21, 1.49	81
	6	19	39	0.87 (0.77, 1.06)	-0.17, 2.41	95
	13	21	42	0.86 (0.67, 1.05)	-0.08, 2.58	95
	27	12	24	0.97 (0.73, 1.21)	0.05, 2.71	95
	52	11	22	0.98 (0.73, 1.23)	0.00, 2.24	93
ODQ	1	5	12	0.72 (0.54, 0.89)	0.41, 1.37	53
	6	14	30	0.86 (0.65, 1.08)	-0.09, 3.40	89
	13	12	24	0.73 (0.56, 0.89)	-0.23, 3.50	87
	27	10	20	0.95 (0.77, 1.12)	0.24, 2.10	83
	52	12	24	0.99 (0.80, 1.19)	-0.05, 2.63	89
	104	4	8	0.93 (0.66, 1.21)	0.39, 1.80	87

^aPooled SMDs (using random effects model) of one arm randomly selected from each trial.

TABLE 3 Pooled SMDs for pain for trial arms stratified by type of treatment

Types of treatments	6 weeks		13 weeks		27 weeks		52 weeks	
	n ^a	SMD ^b						
Index	13	0.75	29	0.91	25	1.01	23	0.85
Active comparator	14	0.82	25	0.87	23	1.02	21	0.89
Placebo, usual care and waiting list ^c	6	0.83	13	0.99	11	1.07	10	0.91
Pharmacological	2	1.00	4	0.77	2	0.61		
Non-pharmacological	20	0.76	53	1.01	49	1.10	47	0.92
Other	5	1.15	5	1.38	5	1.16	3	0.12
Mixed	3	0.63	5	0.70	3	0.75	5	0.68

^aNumber of trial arms. ^bPooled SMDs (using random effects model) of all trial arms. ^cThese trial arms were grouped because of small numbers.

Natural history of low back pain and regression to the mean

The natural history of symptoms and regression to the mean phenomenon could be suggested as a ready and simple explanation. The common pattern of responses could, for a large part, be explained by the natural history of NSLBP. The question is whether it is simply the only explanation or if there are other factors in play given the wide heterogeneity in responses.

Examining the exact natural history (defined as the development of a condition in the absence of treatment [32]) of a disease or symptom, such as NSLBP, is difficult to achieve. However, systematic reviews [33, 34] have attempted to examine the clinical course of NSLBP

(the development of a condition in the presence of treatment [32]) and have shown it to be characterized by a rapid improvement in symptoms within the first 3 months after presentation that becomes more gradual thereafter, which is consistent with our findings.

Compared with participants in trials conducted in health care settings (general practice, occupational therapy departments and physiotherapy departments), participants from the general population trials might not be actively seeking health care at the time of entry into the trials. Symptom improvement among this latter group, therefore, would be nearer to representing the natural history of NSLBP. The larger overall treatment response we found among participants of trials in health care settings might

well be related to the likelihood that people seek health care at a time when pain is worst, leading to larger reductions in symptoms following treatment in these clinical populations.

This would also imply that the magnitude of improvement in symptoms depends, among other factors, on their severity at baseline (more severe symptoms at baseline are associated with larger improvement). The absolute improvements were indeed associated with baseline levels of pain, but the relative rate of improvement remained the same regardless of severity at baseline. Although the natural history or clinical course of NSLBP could contribute to the pattern of symptom improvement shown in the trials included in our review, it is unlikely to be the only explanation.

Regression towards the mean [35] could add an explanation for the large improvement in symptoms early after start of treatment, in particular among those patients with the most severe symptoms at the start of treatment. It is difficult to disentangle these statistical effects from improvement related to the natural history of back pain, with patients being enrolled in trials usually at a stage of more severe symptoms.

Non-specific effects of treatments

Overall response to treatment in clinical trials, cohorts or clinical practice, is not only influenced by the active or specific components of the treatment, but can also be influenced by various other factors. Some of these factors relate to characteristics of the pain problem (as described above), or to the patients themselves [36] such as their beliefs, expectations and experiences with other illnesses, previous episodes of the illness or with previous use of the current treatment or other treatments [37].

It can also be influenced by factors related to the practitioners providing the treatment [38] such as their previous experience with the use of the treatment and their expectation and knowledge of the clinical course of the illness. The actual enrolment in a trial [39], the environment and nature of the communication between the patient and the practitioner [40] and the characteristics of the treatment [41, 42] all represent examples of other influential factors. These examples of factors are by definition non-specific, i.e. not specific to the ingredient of the particular treatment used, but are associated with the process of its provision and reception.

These factors might be at play influencing responses to treatment in clinical trials. The exact size of the influence of these non-specific factors on response to treatment is not known. Whether it is equivalent to the size of the specific effect of the treatment itself or even larger would be interesting to explore. The importance of the influence of these factors is that some of them might explain the common pattern of improvement, such as being enrolled in a trial or the attention given in the trial. On the other hand, some might explain the heterogeneity between responses, such as expectations regarding treatment or aspects of patient-practitioner interaction.

Mean vs individual responses to treatment

Data on responses to treatments in clinical trials, which we have used in this review, were presented in their published reports as aggregated data of responses of their individual participants. Aggregate data tend, by definition, to neutralize or homogenize individual variation and may result in mean responses that would seem more homogeneous than they are at individual level.

Responses of individual participants can vary because of various factors, some of which are related to the nature or severity of symptoms. By definition, we do not know the cause(s) of NSLBP, nor do we always fully understand how treatments may work. Sufferers include a heterogeneous group of patients with a possible variety of underlying causes for their symptoms. It might be the case that by lumping together such a heterogeneous group of patients in trials and presenting their responses as mean scores would inevitably produce non-conclusive outcomes.

Subgroups of patients with certain common characteristics may respond in a different way to treatment than another group with different characteristics. This has led an increasing number of researchers to investigate whether we are targeting the right subgroups of patients with the right treatments [43–45]. This approach is still being developed to establish clear empirical evidence for methods of subgrouping that would be feasible in clinical practice and more accurately predict treatment outcome [46–48].

This also raises an old issue of whether it is better to use aggregate or individual patient data when conducting meta-analyses. The benefit of using individual patient data is obvious: to detect individual differences in responses as well as providing power to any subgroup analyses. However, the great effort and time required can be prohibitive, and therefore it is the systematic reviews and meta-analyses of aggregate data that are most commonly conducted. The outcomes of these reviews provide a major source of evidence on which practicing health practitioners as well as health care policymakers rely in their decision-making. It is not unreasonable to be cautious when judging the ability of these reviews to provide evidence on effectiveness of individual treatments as well as the applicability of their results to the individual patient.

Large overall responses to treatment

An important finding from our review is the large response to treatment common in all trial arms, active as well as placebo, usual care or waiting list arms. It seems ironic that we clearly have evidence for a large overall improvement in back pain symptoms in all arms of clinical trials, while more and more trials are unable to show clear evidence for the effectiveness of the active treatments. Randomized controlled trials are designed to attribute the final benefit of a treatment solely to the difference between its effect and that of the comparator treatment. Such explanatory trials are important and necessary to study the efficacy of new interventions, but it is also important to explore the role of other factors, in addition to

the specific ingredient of the treatment used. Addressing these non-specific factors, which may include factors related to the patients, practitioners, the setting or the way treatments are designed and delivered, may help us better understand responses in clinical trials.

Strengths and limitations

Choice of outcome measures

We chose outcome measures that were most commonly used in order to include a sufficiently large number of trials to examine the evidence. A small number of trials used other measures such as patient's global perceived effect or measures to assess depression or return to work, but these outcome measures did not have standardized definitions or scales which would allow meaningful pooling of the trials.

Hence, although these measures would add to describing the totality of patient experience and provide an important representation of patient response to treatment, we did not include them in this analysis. Recommendations have been made towards a standard group of core outcome measures for use in low back pain trials [49], which would facilitate comparison and combination of results.

Choice of data source

Using CENTRAL database as the source of the included trials satisfied the aims of this review, namely providing us with a large pool of clinical trials on a wide range of primary care treatments for NSLBP to explore the evidence for a common pattern for the overall response to treatments.

Only trials published in English were included. This means that some trials published in other languages will have been missed, but it was not the objective of this review to include all available evidence on a specific intervention. It may be argued that this language restriction could lead to selection bias as non-specific effects of treatments may be influenced by local cultural factors or differences in the delivery or quality of health care. However, there are a number of trials included which were published in English even though they were conducted among non-English-speaking populations. Therefore, the review does include trials from a variety of countries and cultures. There was no other evidence for selection bias as the included trials covered a wide range of treatments in a wide range of primary care settings with no evidence of systematic lack of a particular group or type of trials.

Methodological limitations

Subgroup stratification of trials resulted in small numbers of trials in each subgroup. Any conclusions based on related analyses should, therefore, be made with caution. Although we used a random effects model in the meta-analysis, pooling effect sizes with such a high heterogeneity is still problematic. However, it is important to emphasize that the purpose of pooling for this review was merely to further assess and present the pattern of overall

responses rather than to calculate estimates of treatment effect size, which is usually performed in meta-analyses of effects of particular treatments.

We did not calculate effect sizes based on differences between arms of trials, but responses within each arm. Furthermore, we included trials on a wide variety of disparate types of treatments and therefore even if heterogeneity were to be low, the result of such pooling would not be clinically meaningful. Any conclusions drawn from pooling in our review, therefore, should be made within the context of the particular purpose for using it here.

Conclusions and implications

We have shown that overall responses in NSLBP clinical trials seem to be large and follow a common standard pattern of rapid early improvement followed by a plateau irrespective of treatment. Given such a similar pattern of responses any real effect of treatment may be difficult to detect. It is important to explore factors that influence symptom improvement in clinical trials that apparently happen in active treatment arms as well as in placebo or waiting list arms with only modest differences. Duration of symptoms (acute, subacute or chronic), severity of symptoms at baseline, patients' preference for treatments, patients' expectations and practitioner-patient communications are examples of factors that would influence individual responses to treatment. Identifying such factors would be the first step towards utilizing and harnessing their influence to improve patient outcome in low back pain trials, and this would require further research.

Exploring responses to treatments for pain conditions other than low back pain, and indeed for other medical conditions, was beyond the scope of this review. It would be interesting to explore whether our findings are restricted to NSLBP and therefore reflect in part the poorly understood nature of this condition; or if they could be reproduced in other medical conditions, raising wider issues about how we assess treatment and symptom progression in general.

Rheumatology key messages

- Symptoms of NSLBP improve in trials in a common pattern apparently regardless of the treatments.
- Factors other than the treatments may have a large role in trials and exploring them is important.

Acknowledgements

Funding: This work was undertaken as part of a PhD project funded through an Arthritis Research UK primary care fellowship, fellowship number 17890.

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology Online*.

References

- 1 van Tulder MW, Koes BW, Bombardier C. Low back pain. *Clin Rheumatol* 2002;16:761–75.
- 2 Pengel LHM, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. *Br Med J* 2003;327:323–8.
- 3 Thomas E, Silman AJ, Croft PR, Papageorgiou AC, Jayson MI, Macfarlane GJ. Predicting who develops chronic low back pain in primary care: a prospective study. *Br Med J* 1999;318:1662–7.
- 4 Jellema P, van der Windt DAWM, van der Horst HE, Stalman WAB, Bouter LM. Predictors of an unfavourable course of low back pain in general practice. Comparison of four instruments. *Br J Gen Pract* 2007;57:15–22.
- 5 Andersson GB. Epidemiological features of chronic low back pain. *Lancet* 1999;354:581–5.
- 6 Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain* 2000;84:95–103.
- 7 Dunn KM, Croft PR. Classification of low back pain in primary care: using bothersomeness to identify the most severe cases. *Spine* 2005;30:1887–92.
- 8 Koes BW, Malmivaara A, van Tulder MW. Trend in methodological quality of randomised clinical trials in low back pain. *Clin Rheumatol* 2005;19:529–39.
- 9 van Tulder MW, Koes BW, Malmivaara A. Outcome of non-invasive treatment modalities on back pain: an evidence-based review. *Eur Spine J* 2006;15(Suppl. 1): S64–81.
- 10 Licciardone JC, Brimhall AK, King LN. Osteopathic manipulative treatment for low back pain: a systematic review and meta-analysis of randomized controlled trials. *Musculoskelet Disord* 2005;6:43.
- 11 Keller A, Hayden J, Bombardier C, van Tulder M. Effect sizes of nonsurgical treatments of non-specific low-back pain. *Eur Spine J* 2007;16:1776–88.
- 12 Machado LAC, Kamper SJ, Herbert RD, Maher CG, McAuley JH. Analgesic effects of treatments for non-specific low back pain: a meta-analysis of placebo-controlled randomized trials. *Rheumatology* 2009;5:520.
- 13 Hay EM, Mullis R, Lewis M et al. Comparison of physical treatments versus a brief pain management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet* 2005;365:2024–30.
- 14 Carr JL, Klaber-Moffett JA, Howarth E et al. A randomised trial comparing a group exercise programme for back pain patients with individual physiotherapy in a severely deprived area. *Disabil Rehabil* 2005;27:929–37.
- 15 Frost H, Lamb SE, Doll HA, Carver PT, Stewart-Brown S. Randomised controlled trial of physiotherapy compared with advice for low back pain. *Br Med J* 2004;329:708.
- 16 Jellema P, Van der Windt DAWM, Van der Horst HE, Twisk JWR, Stalman WAB, Bouter LM. Should treatment of (sub) acute low back pain be aimed at psychosocial prognostic factors? Results of a cluster-randomized clinical trial in general practice. *Br Med J* 2005;331:84–8.
- 17 UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *Br Med J* 2004;329:1377–80.
- 18 Leboeuf C, Grbnstvedt A, Borge JA et al. The Nordic back pain subpopulation program: a 1-year prospective multicentre study of outcomes of persistent low-back pain in chiropractic patients. *J Manipul Physiol Therap* 2005;28:90–6.
- 19 Burton AK, Tillotson KM, Main CJ, Hollis S. Psychosocial predictors of outcome in acute and subchronic low back trouble. *Spine* 1995;20:722–8.
- 20 Krismer M, van Tulder M. Low back pain (non-specific). *Clin Rheumatol* 2007;21:77–91.
- 21 Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis* 1987;40: 1129–233.
- 22 Roland M, Morris R. A study of the natural history of back pain. I: development of a reliable and sensitive measure of disability in low back pain. *Spine* 1983;8:141–4.
- 23 Fairbank JCT, Davies JB, Mbaot JC, O'Brien JT. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
- 24 The Cochrane Collaboration Website. www.cochrane.org (4 May 2007, date last accessed).
- 25 van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003;28:1290–9.
- 26 Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005;5:13.
- 27 Egger M, Smith G, Altman DG. Systematic reviews in health care meta-analysis in context. London: BMJ Books, 2001.
- 28 Morris SB. Distribution of the standardised mean change effect size for meta-analysis on repeated measures. *Br J Math Stat Psychol* 2000;53:17–29.
- 29 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Br Med J* 2003;327:557–60.
- 30 Der Simonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
- 31 Cohen J. Statistical power analysis for the behavioural sciences. New York: Academic, 1997.
- 32 von Korff. Studying the natural history of back pain. *Spine* 1994;19:2041S–6S.
- 33 Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. *Eur Spine J* 2003;12:149–65.
- 34 Pengel L, Herbert R, Maher C, Refshauge K. Acute low back pain: systematic review of its prognosis. *Br Med J* 2003;327:323–8.
- 35 Morton V, Torgerson DJ. Regression to the means: treatment effect without the intervention. *J Eval Clin Pract* 2005;11:59–65.
- 36 Price D, Chung K, Robinson M. Conditioning, expectation, and desire for relief in placebo analgesia. *Semin Pain Med* 2005;3:15–21.
- 37 Klinger R, Soost SB, Flor H, Worm M. Classical conditioning and expectancy in placebo hypoalgesia: a randomized controlled study in patients with atopic

- dermatitis and persons with healthy skin. *Pain* 2007;128:31–9.
- 38 Brody H. The doctor as therapeutic agent: a placebo effect research agenda. In: Harrington A, ed. *The placebo effect: an interdisciplinary exploration*. Cambridge, MA: Harvard University Press, 1997:77–92.
- 39 Brauholtz D, Edwards S, Lilford R. Are randomized clinical trials good for us (in the short term)? Evidence for a ‘trial effect’. *J Clin Epidemiol* 2001;54:217–24.
- 40 Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet* 2001;357:757–62.
- 41 Braithwaite A, Cooper P. Analgesic effects of branding in treatment of headaches. *Br Med J* 1981;282:1576–8.
- 42 Kleijnen J, de Craen AJM, van Everdingen JA, Krol L. Placebo effect in double-blind clinical trials: a review of interactions with medications. *Lancet* 1994;344:1347–9.
- 43 Denison E, Asenlöf P, Sandborgh M, Lindberg P. Musculoskeletal pain in primary health care: subgroups based on pain intensity, disability, self-efficacy, and fear-avoidance variables. *J Pain* 2007;8:67–74.
- 44 Childs JD, Fritz JM, Flynn TW et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Intern Med* 2004;141:920–8.
- 45 Boersma K, Linton SJ. Screening to identify patients at risk: profiles of psychological risk factors for early intervention. *Clin J Pain* 2005;21:38–43.
- 46 Schellingerhout JM, Verhagen AP, Heymans MW et al. Which subgroups of patients with non-specific neck pain are more likely to benefit from spinal manipulation therapy, physiotherapy or usual care? *Pain* 2008;139:670–80.
- 47 Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute ‘nonspecific’ low back pain: results of a randomized clinical trial. *Spine* 2006;31:623–31.
- 48 Hay EM, Dunn KM, Hill JC et al. A randomised clinical trial of sub-grouping and targeted treatment for low back pain compared with best current care. The StarT Back Trial Study Protocol. *BMC Musculoskelet Disord* 2008;9:58.
- 49 Dworkin R, Turk D, Farrar J et al. Core outcome measures for chronic pain clinical trials: IMPACT recommendations. *Pain* 2005;113:9–19.