Do baseline characteristics predict response to treatment for low back pain? Secondary analysis of the UK BEAM dataset [ISRCTN32683578]

M. R. Underwood[1], V. Morton[2] and A. Farrin[3] on behalf of the UK BEAM trial team[2,4]

Objectives. To identify characteristics of randomized controlled trial participants which predict greater benefits from physical treatments for low back pain. If successful, this would allow more appropriate selection of patients for different treatments.

Methods. We did a secondary analysis of the UK Back pain Exercise And Manipulation trial (UK BEAM $n = 1334$) dataset to identify baseline characteristics predicting response to manipulation, exercise and manipulation followed by exercise (combined treatment). Rather than simply identifying factors associated with overall outcome, we tested for the statistical significance of the interaction between treatment allocation, baseline characteristics and outcome to identify factors that predicted response to treatment. We also did a *post-hoc* subgroup analysis to present separate results for trial participants with subacute and chronic low back pain to inform future evidence synthesis.

Results. Age, work status, age of leaving school, 'pain and disability', 'quality of life' and 'beliefs' at baseline all predicted overall outcome. None of these predicted response to treatment. In those allocated to combined treatment, there was a suggestion that expecting treatment to be helpful might improve outcome at 1 yr. Episode length at study entry did not predict response to treatment.

Conclusion. Baseline participant characteristics did not predict response to the UK BEAM treatment packages. Using recognized prognostic variables to select patients for different treatment packages, without first demonstrating that these factors affect response to treatment, may be inappropriate. In particular, this analysis suggests that the distinction between subacute and chronic low back pain may not be useful when considering treatment choices.

KEY WORDS: Back Pain, Randomised Controlled Trial, Prognosis.

Introduction

Low back pain has a major health and social impact [1]. There are now many randomized controlled trials of different treatment approaches for low back pain. However, these rarely show more than a small to moderate overall benefit from the treatment [2]. The need to identify which subgroups of back pain patients are likely to gain the most from different interventions is well recognized [3]. There is substantial literature on the predictors of poor outcome from low back pain. Only a few studies have considered, within randomized controlled trials, which patient factors identify those who are most likely to gain the greatest benefit from the treatments tested.

Karjalainen [4] found that a brief intervention was most effective in those who perceived that they had greatest risk of not recovering. Hagen [5] found that the benefit of attending a spine clinic was greatest in those with a low-health locus of control at 3 months and at 1 yr; and in those with constant pain at 3 months but not at 1 yr. Hagen also found that presence of gastro-intestinal complaints predicted who was most likely to respond to treatment. Both Karjalainen and Hagen looked for an interaction between baseline characteristics and treatment group in a multivariate model to identify possible modifiers of treatment effect.

In *post-hoc* subgroup analyses of randomized controlled trials, Burton [6] found that an educational booklet had had a beneficial effect only in those with high fear-avoidance scores; and Klaber Moffett [7] found that those with high fear avoidance benefited most from an exercise programme. George *et al.* [8] found that patients with high initial fear-avoidance beliefs benefited more from fear-avoidance-based physiotherapy than those with low fear-avoidance beliefs. Childs *et al.* [9, 10] in a randomized controlled trial of exercise compared with exercise plus manipulation, used a clinical prediction rule (presence of four of the following: short symptom duration, fear-avoidance beliefs, lumbar hypomobility, hip internal rotation range of motion and no symptoms distal to the knee) that identified those with a good prognosis and found that those allocated to manipulation, who were positive on the prediction rule, gained the greatest benefit.

We have completed a large, primary care-based, randomized controlled trial of physical treatments for back pain: the UK Back pain Exercise And Manipulation trial (UK BEAM) [11–13]. Our principal findings were that when compared with 'best care' in general practice [14, 15], a package of spinal manipulation [16] produced a small to moderate benefit at 3 months and a small benefit at 1 yr; that a programme of exercise [17] produced a small benefit at 3 months, but not 12 months; and that manipulation followed by exercise (combined treatment) produced a moderate benefit at 3 months and a small benefit at 1 yr.

Our primary outcome measure was the Roland Morris Disability Questionnaire (RMDQ) [18]. The largest additional benefit from adding physical treatments to 'best care' in general practice was 1.87 [95% confidence interval (CI) 1.15–2.60] RMDQ points, for combined treatment at 3 months following randomization. For our sample size calculation, we set the clinically important difference at 2.5 (S.D. 4.0) RMDQ points [13]. The mean benefit we observed is unlikely to be important for an individual patient [19].

UK BEAM included participants with 'simple low back pain' lasting at least 4 weeks and randomized them to four different treatment packages. Within each of these packages, the individual general practitioner or therapist had considerable freedom to decide how he/she treated individual patients. The diagnosis of simple low back pain reflects our inability to identify meaningful definitions of different back pain subgroups [3]. Thus UK BEAM

[1]Centre for Health Sciences, Barts and The London, Queen Mary's School of Medicine and Dentistry, [2]York Trials Unit, Department of Health Sciences, University of York, [3]Clinical Trials Research Unit, University of Leeds and [4]MRC General Practice Research Framework, UK.

Correspondence to: Martin Underwood, Barts and The London, Centre for Health Sciences, Abernethy Building, 2 Newark Street, Whitechapel, London E1 2AT, UK. E-mail: m.underwood@qmul.ac.uk

recruited a heterogeneous group of patients with low back pain. Nevertheless, we observed statistically significant benefits from our additional treatment packages. Some participants gained an important benefit from some of the additional treatments; for some participants they were ineffective; and some participants might have been made worse. Qualitative analysis of participants' comments supports the notion that some gained substantial benefits from the intervention packages [20]. Identifying which patients are more likely to benefit from these treatment packages would allow more efficient targeting of resources.

Duration of the pain episode is commonly used to inform the management of low back pain. Current guidelines for the management of back pain generally consider acute/subacute (<3 months) and chronic (>3 months) pain separately. Our primary analysis, which pooled patients with both subacute and chronic pain, does not allow our results directly to inform guideline development because meta-analysts and guideline developers typically exclude trials including a mixture of participants with a long and short duration of low back pain. The main results from UK BEAM led us to hypothesize that, as well as duration, other baseline characteristics could differentially affect the size of benefits from treatments. For example, those allocated to exercise, unlike those randomized to manipulation, demonstrated large changes in fear-avoidance beliefs. This suggests that people with high fear-avoidance might gain the greatest benefit from exercise and those with low fear-avoidance the greatest benefit from manipulation.

In summary, little is known about the relationships between patient characteristics and response to specific low back pain treatments. If identifiable individual patient factors predict treatment response, back pain treatment could be substantially improved by targeting interventions at those likely to gain the greatest benefit. We have conducted further analyses of the UK BEAM data to assess the impact of participant baseline characteristics on response to treatment and to present appropriate data to inform future meta-analyses and guideline development.

Methods

UK BEAM is reported in detail elsewhere [11–13]. We briefly summarize the method here.

Participant recruitment and follow-up

We recruited 1334 participants from 181 general practices from the Medical Research Council General Practice Research Framework (www.gprf-mrc.ac.uk) in 14 clusters across the UK. Participants were patients aged 18–64 who had consulted these practices with simple low back pain that failed to resolve after their consultation. All participants' current episode of back pain had lasted for at least 4 weeks. Before randomization, research nurses based in each practice collected routine demographic data and participants completed a baseline assessment questionnaire. Follow-up was by postal questionnaire 3 months and 1 yr after randomization. Ethical review was provided by Northern and Yorkshire Multi-Centre Research Ethics Committee.

Interventions: 'best care' in general practice—the 'control' treatment

We trained all general practice teams in the 'active management' of back pain [14, 15] and they provided patients with copies of 'The Back Book' [21].

Exercise programme. This consisted of an initial individual assessment followed by up to nine group classes over 12 weeks delivered in community facilities [17].

Spinal manipulation package. The UK chiropractic, osteopathic and physiotherapy professions agreed a package of techniques developed by a multidisciplinary group [16]. We invited participants to attend up to eight sessions over up to 12 weeks.

Combined treatment. Participants allocated to this group received up to 6 weeks of manipulation followed by up to 6 weeks of exercise. Apart from the timing, all aspects of treatment were identical to those in the manipulation-only or exercise-only groups.

Randomization

We randomized participants between four treatment options—'best care in general practice', 'best care' plus spinal manipulation package, 'best care' plus class-based exercise program and 'best care' plus combined treatment.

Baseline data collection

Before randomization, we collected the following baseline data.

Health status. The RMDQ [18], a well-established outcome measure for community-based back pain studies [22], was our primary outcome measure. We measured generic health status using the Short Form 36-item Health Survey (SF-36) [23], and the EQ5D [24]. We complemented these with measures for back pain and disability derived from the chronic pain grade (MVK pain and MVK disability) [25, 26] and single-item questions on 'troublesomeness' of low back pain and leg pain anglicized from previously used 'bothersomeness' questions [27]. These questions were on a 5-point Likert scale, ranging from not at all troublesome to extremely troublesome. These were converted into 1–5 numerical values for this analysis.

Back pain beliefs. We measured beliefs about back pain using the physical sub-scale of the Fear Avoidance Beliefs Questionnaire physical (FABQ) [28] and the inevitability subscale of the Back Beliefs Questionnaire (BBQ) [29].

Psychological instruments. We assessed psychological state using the Distress and Risk Assessment Method (DRAM) which is derived from Modified Somatic Perception Questionnaire and Modified Zung Depression Inventory [30].

Treatment expectations. We asked participants before randomization for their expectation for a benefit from each treatment option using a 3-point Likert scale: how helpful do you think treatment from (each treatment) will be for your back pain, very helpful, helpful or not helpful?

Other data. We collected basic demographic data, information on educational attainment, work status and duration of current episode of back pain.

Statistical methods

In our baseline dataset, we had 18 candidate explanatory variables: 11 health status measures, four demographic measures, duration of episode, expectation of benefit from treatment and randomized treatment (Table 1). In an initial exploratory analysis, we found that our 11 health status measures were highly correlated with each other. This would cause problems of multicollinearity in a multiple regression model, making the results difficult to interpret. To reduce the number of variables we put into the regression model, we first carried out a principal component analysis using varimax rotation with Kaiser normalization to determine a smaller number of uncorrelated components that could then be used in the regression

TABLE 1. Candidate explanatory baseline variables

Health status measures
  Roland Morris disability questionnaire [18]
  SF-36 mental component score (MCS) [23]
  SF-36 physical component score (PCS) [23]
  EQ-5D [24]
  Modified Von Korff—disability [25, 26]
  Modified Von Korff—pain [25, 26]
  Troublesomeness questions—low back pain [27]
  Troublesomeness questions—leg pain [27]
  Fear avoidance beliefs questionnaire physical subscale [28]
  Back beliefs questionnaire [29]
  Modified somatic perception questionnaire [30]
  Modified Zung [30]
Demographic variables
  Age
  Sex
  Age of leaving full-time education
  Working/not working
Duration
  Length of current episode in days
Expectation for helpfulness
  Response to pre-randomisation question of whether they thought their eventual
    treatment allocation would be helpful
Treatment allocation

TABLE 2. Rotated component matrix baseline data ($N = 1158$)

	Pain and disability	Quality of life	Beliefs
RMDQ	0.580	0.219	0.398
SF-36 MCS score (*T*-score)		−0.895	−0.111
SF-36 PCS score (*T*-score)	−0.730	0.148	−0.395
EQ 5D	−0.615	−0.347	−0.226
Modified Von Korff—pain	0.743	0.244	0.234
Modified Von Korff—disability	0.652	0.251	0.408
Troublesomeness—low back pain	0.744	0.186	
Troublesomeness—leg pain	0.630	0.148	−0.125
FABQ physical—baseline	0.117		0.836
BBQ baseline	−0.167	−0.284	−0.710
MSPQ	0.261	0.725	
Modified Zung	0.247	0.827	0.203

Extraction method: principal component analysis. Rotation method: varimax with Kaiser Normalization. Rotation converged in five iterations.

analyses [31]. Principal component analysis produces a set of Eigen values and a corresponding Eigen vector; the Eigen vectors contain weights that can be used to compute the new variables from the original ones. Each of the new variables is a combination of the weighted original variables with the weighting determined by the Eigen vectors for each of the new 'factors' [32].

We then carried out initial analyses, using our new variables ('Pain and Disability', 'Quality of Life' and 'Beliefs'), to explore the extent to which all of our baseline variables predicted overall outcome at 3 and 12 months, respectively. We did an ordinary least square regression, with the RMDQ score at 3 or 12 months as the dependent variable. To test the hypothesis that patient characteristics could predict response to different treatments, we analysed our 3- and 12-month outcomes testing for the interaction between treatment allocation and each of our baseline variables [33].

This approach makes the best use of the available data and avoids the problems inherent in doing multiple subgroup analyses [34, 35]. For these analyses we excluded data from participants recruited from the 13 pilot study practices, who were included in our main analyses, because of slight differences in the coding of the 'helpfulness' question in these participants.

Because of the importance ascribed to episode duration in planning treatment services, we proceeded to carry out a further analysis looking for an interaction in a model that included only duration and treatment allocation. This analysis again avoids problems with *post-hoc* subgroup analyses. Analysis of covariance, adjusting for baseline RMDQ score, was used to investigate any potential relationships between treatment allocation and the binary duration variable (<90 days and ≥90 days). The variable of interest was the interaction between the binary duration variable and treatment allocation.

Notwithstanding our reservations about *post-hoc* subgroup analyses, because it is currently common practice to use a defined cut-off time for episode duration when synthesizing evidence on the treatment of low back pain, we carried out a *post-hoc* subgroup analysis presenting the results for those with <90 days and ≥90 days separately to inform future evidence synthesis.

Results

Baseline data

We had complete datasets on 1116 participants at baseline that contributed to this analysis. We identified three factors with an

TABLE 3. Multivariate analysis of overall predictors of outcome

	Three months ($n = 723$)			One year ($n = 700$)		
	β	P	95% CI	β	P	95% CI
Manip+ex[a]	−1.8	<0.001	(−2.8, −0.9)	−1.4	0.005	(−2.3, −0.4)
Manipulation[a]	−1.3	0.003	(−2.2, −0.4)	−0.6	0.178	(−1.8−0.1)
Exercise[a]	−1.3	0.005	(−2.2, −0.4)	−0.8	0.178	(−1.5−0.3)
Additional education[b]	−0.7	0.033	(−1.3, −0.1)	−1.1	0.001	(−1.7, −0.4)
Working[c]	−1.3	0.001	(−2.0, −0.5)	−1.3	0.001	(−2.1, −0.6)
Age[d]	0.05	0.001	(0.02, −0.08)	0.1	0.002	(0.02, 0.08)
Very helpful[e]	−0.6	0.202	(−1.6, −0.4)	−0.6	0.272	(−1.6, 0.5)
Helpful[e]	−0.1	0.776	(−1.0, −0.8)	0.2	0.722	(−0.8, 1.1)
Male[f]	−0.7	0.037	(−1.3, −0.04)	0.6	0.081	(−0.1, 1.2)
Pain and disability[g]	1.7	<0.001	(1.4, 2.0)	2.0	<0.001	(1.7, 2.3)
Quality of life[h]	0.7	<0.001	(0.4, 1.1)	0.8	<0.001	(0.5, 1.1)
Beliefs[i]	1.2	<0.001	(0.9, 1.5)	1.0	<0.001	(0.7, 1.4)
Episode length[j]	0.4	<0.001	(0.2, 0.6)	0.5	<0.001	(0.3, 0.7)

Parameter estimate for each factor, corrected for all variables in this model. Negative values indicate more improvement. Outcome = RMDQ score.
[a]*vs* 'best care'.
[b]*vs* no additional education after age 16.
[c]*vs* not working.
[d]For each year age difference.
[e]Allocated treatment not helpful.
[f]*vs* female.
[g]Pain and disability—higher score represents a greater degree of pain and disability (worse).
[h]Quality of life—a higher score represents poorer quality of life.
[i]Beliefs—a higher score represents less favourable beliefs.
[j]Subacute *vs* chronic.

Eigen Value of >1; 'Pain and Disability', 'Quality of Life' and 'Beliefs'. Together these factors explained 64.1% of the variance in the baseline data: 42.5, 12.7 and 8.9%, respectively (Table 2). We then used the individual participants' scores for each factor in the next stage of the analysis. Participants had high expectations for success from the different treatment options: 92, 86 and 97%, respectively, expected manipulation, exercise, or combined treatment to be helpful or very helpful, though only 63% had the same level of expectation for general practice care.

Overall predictors of outcome

The mean improvement in participants randomized to any of the active interventions was greater than usual care at 3 months. This was sustained at 1 yr for manipulation and combined treatment but not for exercise. Higher levels of education, lower age, shorter episode length and being at work were all associated with a better outcome. Higher values in our three new factors, that is more pain and disability, poorer quality of life and less favourable beliefs, were all strongly associated with a poorer outcome at both 3 months and 1 yr (Table 3). Our baseline variables predicted

TABLE 4. Interaction between significant baseline variables and treatment allocation

		Month 3			Month 12		
		β	P	95% CI	β	P	95% CI
Combined	Episode length	−0.3	0.302	(−0.80, 0.25)	0.0	0.950	(−0.56, 0.53)
Combined	Beliefs	−0.1	0.905	(−0.90, 0.80)	−0.3	0.474	(−1.20, 0.56)
Combined	Quality of life	0.6	0.174	(−0.26, 1.43)	0.5	0.252	(−0.36, 1.38)
Combined	Pain/disability	0.2	0.745	(−0.73, 1.02)	0.2	0.703	(−0.72, 1.07)
Combined	Gender	0.3	0.736	(−1.41, 2.00)	−0.4	0.639	(−2.19, 1.34)
Combined	Treatment, helpful	−3.2	0.073	(−6.74, 0.30)	−3.8	0.038	(−7.39, −0.20)*
Combined	Treatment, very helpful	−2.2	0.192	(−5.49, 1.11)	−4.0	0.019	(−7.38, −0.67)*
Combined	Age	0.0	0.734	(−0.09, 0.07)	0.0	0.429	(−0.113, 0.048)
Combined	Work	−0.2	0.813	(−2.22, 1.74)	−0.6	0.548	(−2.67, 1.42)
Combined	Education	0.0	0.963	(−1.78, 1.69)	0.8	0.411	(−1.04, 2.54)
Exercise	Episode length	−0.1	0.840	(−0.96, 0.78)	0.0	0.998	(−0.88, 0.88)
Exercise	Beliefs	−0.1	0.793	(−0.99, 0.76)	0.3	0.582	(−0.66, 1.18)
Exercise	Quality of life	−0.2	0.706	(−1.04, 0.71)	0.0	0.997	(−0.90, 0.91)
Exercise	Pain/disability	0.2	0.816	(−1.55, 1.97)	−1.0	0.312	(−2.80, 0.90)
Exercise	Gender	−1.2	0.384	(−3.96, 1.53)	−1.3	0.400	(−4.17, 1.67)
Exercise	Treatment, helpful	0.4	0.724	(−1.92, 2.75)	−0.1	0.924	(−2.57, 2.33)
Exercise	Treatment, very helpful	0.0	0.781	(−0.07, 0.10)	0.0	0.427	(−0.13, 0.05)
Exercise	Age	0.7	0.513	(−1.39, 2.78)	−0.9	0.403	(−3.05, 1.23)
Exercise	Work	1.0	0.295	(−0.83, 2.72)	0.7	0.473	(−1.18, 2.54)
Exercise	Education	−0.3	0.347	(−0.76, 0.27)	−0.2	0.546	(−0.71, 0.37)
Manipulation	Episode length	−0.2	0.729	(−1.01, 0.71)	0.1	0.801	(−0.78, 1.00)
Manipulation	Beliefs	−0.8	0.070	(−1.62, 0.06)	−0.4	0.328	(−1.29, 0.43)
Manipulation	Quality of life	1.4	0.118	(−0.35, 3.07)	0.5	0.583	(−1.28, 2.28)
Manipulation	Pain/disability	−1.9	0.176	(−4.61, 0.85)	−2.2	0.143	(−5.16, 0.75)
Manipulation	Gender	0.3	0.788	(−2.06, 2.71)	0.4	0.774	(−2.23, 3.00)
Manipulation	Treatment, helpful	0.0	0.669	(−0.10, 0.06)	−0.1	0.083	(−0.16, 0.01)
Manipulation	Treatment, very helpful	1.6	0.113	(−0.38, 3.60)	1.2	0.250	(−0.86, 3.29)
Manipulation	Age	0.2	0.806	(−1.49, 1.91)	1.1	0.218	(−0.66, 2.90)
Manipulation	Work	−0.3	0.302	(−0.80, 0.25)	0.0	0.950	(−0.56, 0.53)
Manipulation	Education	−0.1	0.905	(−0.90, 0.80)	−0.3	0.474	(−1.20, 0.56)

Parameter estimate for each factor, corrected for all variables in this model. Negative values represent a lower score at follow-up point (better outcome). Outcome = RMDQ score. *p<0.05

TABLE 5. Additional benefit in RMDQ from adding treatment packages to 'best care'

	Three months				One year			
	n	Difference	P	95% CI	n	Difference	P	95% CI
Chronic n=878								
'Best care'	165	–	–	–	157	–	–	–
Combined	172	−1.39	0.043	(−2.74, −0.04)	172	−1.52	0.04	(−2.96, −0.09)
Manipulation	183	−0.91	0.159	(−2.18, 0.36)	175	−0.55	0.44	(−1.92, 0.83)
Exercise	147	−1.24	0.072	(−2.59, 0.11)	142	−1.33	0.08	(−2.79, 0.14)
Sub-acute n=393								
'Best Care'	80	–	–	–	81	–	–	–
Combined	71	−2.06	<0.001	(−2.96, −1.16)	70	−1.15	0.02	(−2.12, −0.18)
Manipulation	90	−1.69	<0.001	(−2.57, 0.80)	83	−1.24	0.01	(−2.21, −0.27)
Exercise	70	−1.23	0.010	(−2.17, −0.29)	65	0.23	0.66	(−0.79, 1.25)

Sixty-three participants did not have duration recorded.

32.7 and 33.2% of the variance in outcome at 3 months and 1 yr, respectively (adjusted R^2).

Interaction between baseline variables and treatment response

The only interactions that approached statistical significance were: (i) between combined treatment and expectation that the treatment would be helpful at baseline and the 1-yr RMDQ score [−4.0 RMDQ points ($P = 0.019$, 95% CI −7.4, −0.7)]; (ii) between combined treatment and expectation that the treatment would be very helpful at baseline and the 1-yr RMDQ score [−3.8 RMDQ points ($P = 0.038$, 95% CI −7.4, −0.2)] (Table 4).

These results suggest that participants allocated to combined treatment who expected this treatment to be helpful or very helpful respectively gain an additional 4.0 and 3.8 points improvement in the RMDQ at 1 yr when compared with those who did not think the treatment would be helpful. These apparently statistically significant results may be chance findings because of the large number of comparisons.

Effect of duration at baseline

There were no statistically significant interactions between episode length as a continuous variable or as a categorical variable (<90 days *vs* ≥90 days) with outcome at either 3 months or 1 yr. The benefits from treatment for the three active interventions for subgroups with subacute and chronic pain are presented in Table 5.

Discussion

Although a number of our baseline measures, age, work status, age of leaving school, 'pain and disability', 'quality of life' and 'beliefs' at baseline, predicted overall outcome, none of these predicted response to treatment. There was a suggestion that expecting treatment to be helpful might improve outcome at 1 yr in those allocated to combined treatment. The duration of the current episode of back pain did not predict whether participants benefited from the treatment packages.

Identifying subgroups

As with all secondary analyses, our findings need to be interpreted with caution. Some of the treatment-effect sizes in this analysis are slightly different from those in the main UK BEAM analysis because of the use of different baseline covariates. The main UK BEAM clinical outcomes paper remains the definitive analysis of the treatment effects. UK BEAM is an unusual randomized controlled trial of back pain because it is of sufficient size to enable us to analyse baseline factors that might predict response to treatment. However, UK BEAM was not designed to identify these predictors, so it is possible that we still had insufficient statistical power to identify such clinically important interactions if they did exist [35].

Our initial factor analysis developed three independent factors from the baseline variables, which equate to broad domains that would be widely accepted as predicting outcome for patients with low back pain ('Pain and Disability', 'Quality of Life' and 'Beliefs'). They all have highly significant associations with outcome, suggesting that they are appropriate factors to use in these analyses. We were not able to derive our new factors in a different dataset than that used for the final analysis; the dataset was not large enough to allow us to derive our factors in one half and test them in the other half. This may lead to us overestimating the importance of these factors. There are inherent weaknesses in pooling multiple outcome instruments that are measuring different constructs. However, only by doing this were we able to test whether these broad factors of 'pain and disability', 'quality of life' and 'beliefs', which resonate for individual clinicians, do have an impact on treatment response.

This article is not focused on identifying overall predictors of outcome for patients with back pain. Nevertheless, it is reassuring that the independent predictors of outcome broadly concur with those found in epidemiological studies. It is disappointing that we were unable to identify any convincing predictors of response to treatment.

Only for expectations of treatment helpfulness in those allocated to combined treatment was there a significant interaction between allocation and response to treatment. In view of the large number of comparisons made, we cannot conclude that these findings can be used to inform treatment choices.

Our baseline measures together explained only 33% of the variance in outcome. Although there are statistically strong associations between our baseline measures and final outcome, two-thirds of what predicts outcome has not been captured by them.

Simply presenting the statistical significance of an association between baseline data and outcome may be misleading as it does not show the actual strength of the association, nor does it equate to positive and negative predictive values. Since we can explain only a small proportion of the overall variance, it is not surprising that even in the large UK BEAM dataset we were unable to predict response to treatment.

This is in marked contrast to other, much smaller studies which, using subgroup analyses, have reported effect sizes for an additional treatment benefit of a similar magnitude to our main treatment effects using different baseline measures [6, 7, 9, 10]. However, it should be noted that both study populations and baseline measures collected differed between these studies and UK BEAM. For example, Childs' clinical prediction rule includes duration of <16 days and results of a physical examination [9]. In contrast, UK BEAM only included people with continuing back pain at least 4 weeks after an initial consultation and did not include a physical examination.

Our analyses suggest that it is inappropriate to use overall predictors of outcome to inform specific treatment choices without first demonstrating that these do indeed predict response to treatment. More encouragingly both Karjalainen and Hagen [4, 5], using a similar statistical approach to the one we

have used here, albeit in different populations and using different baseline measures, have found possible, easily assessed predictors of response to specific treatments. Furthermore, Brennan et al. [36] found that selecting different physical therapy approaches to match participants' baseline characteristics appears to improve outcome. However, a no-treatment or usual-care arm was not included in this study. Further trials that are powered to show the interaction between pre-specified baseline characteristics and treatment outcome, with a usual care control group, are a possibility. Based on current work it would be possible to identify some factors that could be used in such studies. However, such trials are likely to be expensive and time-consuming. For example, the sample size estimate needs to be increased 4-fold to show an interaction of a similar magnitude to the main effect sought [35].

Effect of episode duration

Notwithstanding our reservations about subgroup analyses, there is some utility in considering the distinction between subacute and chronic pain. Contrary to our expectations, episode duration did not predict a differential response to different treatments for patients with subacute or chronic back pain. This study does not allow us to comment on any differential treatment effects for those with acute back pain. Current orthodoxy, endorsed by numerous national and international guidelines, is to consider acute and chronic back pain separately when planning treatment recommendations. Some have challenged this, positing that almost all back pain is a 'chronic problem with an untidy pattern of grumbling symptoms' [37].

Nevertheless, the distinction between acute and chronic low back pain would have utility if it allowed us to target treatment resources more effectively. Our findings do not support the notion that those with less persistent pain gain a greater benefit from treatment; rather, these physical treatments should be equally accessible to all patients with sub-acute or chronic low back.

Our findings support the view that splitting back pain into sub-acute or chronic categories based on the duration of pain alone may be of limited value and that this artificial distinction should be laid to rest. It is not worth trying to draw any inferences from the apparent differences in treatment response according to whether the pain has been present for more or less than 90 days. These results should however be useful for future meta-analyses and guidelines that may continue to draw a distinction between acute/subacute and chronic low back pain.

Conclusions

In the main UK BEAM paper, we argued that although the average effect sizes from our interventions were small to moderate, some patients were likely to have had larger benefits; and that because the interventions were very cost-effective these treatments should be made more generally available. That we have been unable to identify which patients may get the greatest benefit from a particular intervention appears to strengthen that position.

In common with previous trials, for example trials of chiropractic treatment [38, 39], exercise [40] or physiotherapy [41], UK BEAM has provided only a partial answer to the question of how back pain should be treated. The main analyses have shown incontrovertible evidence of a small to moderate short-term benefit from all three intervention packages, when added to 'best care' in general practice, along with small sustained benefits from the manipulation and combined treatment packages. However, the data have failed to identify which patients will gain the greatest benefits from which treatment, or which components of the treatment are responsible for these effects. Since we have already shown that we can achieve these small to moderate benefits at a modest cost, it seems reasonable to continue to advise making these treatments available to patients who are not

spontaneously improving a few weeks after consulting for an episode of back pain of any duration.

Rheumatology key messages

- Participant baseline characteristics did not identify who was more likely to respond to physical treatments for low back pain.
- Duration of episode does not predict which patients will get the greatest benefit from these physical treatments for low back pain.

The authors have declared no conflicts of interest.

References

1 Maniadakis N, Gray A. The economic burden of back pain in the UK. Pain 2000;84:95–103.

2 Wyatt M, Underwood MR, Scheel IB, Cassidy JD, Nagel P. Back pain and health policy research: the what, why, how, who, and when. Spine 2004;29:E468–75.

3 Borkan JM, Koes B, Reis S, Cherkin DC. A report from the Second International Forum for Primary Care Research on Low Back Pain. Reexamining priorities. Spine 1998;23:1992–6.

4 Karjalainen K, Malmivaara A, Mutanen P, Roine R, Hurri H, Pohjolainen T. Mini-intervention for subacute low back pain: two-year follow-up and modifiers of effectiveness. Spine 2004;29:1069–76.

5 Hagen EM, Svensen E, Eriksen HR. Predictors and modifiers of treatment effect influencing sick leave in subacute low back pain patients. Spine 2005;30:2717–23.

6 Burton AK, Waddell G, Tillotson KM, Summerton N. Information and advice to patients with back pain can have a positive effect. A randomized controlled trial of a novel educational booklet in primary care. Spine 1999;24:2484–91.

7 Klaber Moffett JA, Carr J, Howarth E. High fear-avoiders of physical activity benefit from an exercise program for patients with back pain. Spine 2004;29:1167–72.

8 George SZ, Fritz JM, Bialosky JE, Donald DA. The effect of a fear-avoidance-based physical therapy intervention for patients with acute low back pain: results of a randomized clinical trial. Spine 2003;28:2551–60.

9 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.

10 Flynn T, Fritz J, Whitman J et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 2002;27:2835–43.

11 The 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.

12 The UK BEAM Trial Team, United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: cost effectiveness of physical treatments for back pain in primary care. Br Med J 2004;329:1381.

13 The UK BEAM Trial Team, UK Back pain Exercise And Manipulation (UK BEAM) trial - national randomised trial of physical treatments for back pain in primary care: objectives, design and interventions [ISRCTN32683578]. BMC Health Serv Res 2003;3:16.

14 Underwood M, O'Meara S, Harvey E. The acceptability to primary care staff of a multidisciplinary training package on acute back pain guidelines. Fam Pract 2002;19:511–5.

15 Harvey E, Farrin A, Underwood M, Morton V. What's it got to do with us? Including support staff in general practice training sessions. Educ Primary Care 2004;15:606–10.

16 Harvey E, Burton AK, Moffett JK, Breen A. Spinal manipulation for low-back pain: a treatment package agreed to by the UK chiropractic, osteopathy and physiotherapy professional associations. Man Ther 2003;8:46–51.

17 Klaber Moffett J, Frost H. Back to Fitness Programme. Physiotherapy 2000;86:295–305.

18 Roland M, Morris R. 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 1983;8:141–4.

19 Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. J Rheumatol 2001;28:431–8.

20 Underwood MR, Harding G, Klaber MJ. Patient perceptions of physical therapy within a trial for back pain treatments (UK BEAM) [ISRCTN32683578]. Rheumatology (Oxford) 2006;45:751–6.

21 Roland M, Waddell G, Klaber Moffett J, Burton AK, Main CJ, Cantrell T. The Back Book. Norwich: The Stationery Office, 1996.

22 Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 2000;25:3115–24.

23 Ware JEJ, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.

24 The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.

25 Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50:133–49.

26 Underwood MR, Barnett AG, Vickers MR. Evaluation of two time-specific back pain outcome measures. Spine 1999;24:1104–12.

27 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care 1989;27: (3 Suppl):S217–32.

28 Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52:157–68.

29 Symonds TL, Burton AK, Tillotson KM, Main CJ. Do attitudes and beliefs influence work loss due to low back trouble? Occup Med 1996;46:25–32.

30 Main CJ, Wood PL, Hollis S, Spanswick CC, Waddell G. The distress and risk assessment method. A simple patient classification to identify distress and evaluate the risk of poor outcome. Spine 1992;17:42–52.

31 Miles J, Shelvin M. Applying regression and correlation - a guide for students and researchers. Sage, 2001:132, London.

32 Cohen J, Cohen P, West SG, Aiken P. Applied multiple regression/correlation analysis for the behavioural sciences Lawrence Erlbaum Associates, 2003, Hillsdale, NJ.

33 Pocock SJ. Clinical trials: a practical approach, Wiley, 1983 Chichester, UK.

34 Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005;365:176–86.

35 Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229–36.

36 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.

37 Croft PR, MacFarlane GJ, Papageorgiou AC, Thomas E, Silman AJ. Outcome of low back pain in general practice: a prospective study. Br Med J 1998;316:1356–9.

38 Meade TW, Dyer S, Browne W, Townsend J, Frank AO. Low back pain of mechanical origin: randomised comparison of chiropractic and hospital outpatient treatment. Br Med J 1990;300:1431–7.

39 Meade TW, Dyer S, Browne W, Frank AO. Randomised comparison of chiropractic and hospital outpatient management for low back pain: results from extended follow up. Br Med J 1995;311:349–51.

40 Moffett JK, Torgerson D, Bell-Syer S et al. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. Br Med J 1999;319:279–83.

41 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.