Evaluation of a Multi-disciplinary Back Pain Rehabilitation Programme—Individual and Group Perspectives

Andrew Baird · Lisa Worral · Cheryl Haslam ·
Roger Haslam

Abstract

Objective To evaluate the impact of a multi-disciplinary back pain rehabilitation programme using a combination of individual and group change data.

Methods A total of 261 consecutive patients attending an assessment session for the back pain rehabilitation programme completed the SF-36 health survey questionnaire. The patients were requested to complete the questionnaires again at programme completion and at the 6-month follow-up. The Reliable Change Index was used to define 'clinical significance' in terms of the assessment of individual change.

Results Half of those patients considered to be suitable for the programme subsequently completed it. In group terms, non-completers scored lower than completers on all SF-36 scales. Statistically significant improvements were evident for those completing the programme (all scales at $P < 0.000$), with improvement maintained at follow-up. In individual terms, 'clinical significance' was exceeded most frequently in the Physical Functioning and Role Physical scales. Whilst some participants lost previous improvements between completion and follow-up, others improved over this same time period. The majority of those completing the programme showed improvement in at least one scale.

Conclusions Adding assessment of individual change to traditional group change measures provides greater insight into the impact a rehabilitation programme has upon participants' quality of life. Whilst the programme is clearly effective for those who complete it, work is required to limit post-programme deterioration and improve uptake.

Keywords Clinically significant change ·
Low back pain · Rehabilitation · Reliable Change Index

Abbreviations

BP	Bodily pain
CS-BP	Clinically significant improvement in bodily pain
CS-PF	Clinically significant improvement in physical functioning
CS-RP	Clinically significant improvement in role physical
GH	General health
MH	Mental health
PF	Physical functioning
RE	Role emotional
RP	Role physical
SF	Social functioning
VT	Vitality

A. Baird (✉)
Centre for Psychological Research, University of Derby,
Kedleston Road, Derby DE22 1GB, UK
e-mail: A.Baird@derby.ac.uk

L. Worral
Nottingham University Hospitals NHS Trust, Nottingham, UK

C. Haslam · R. Haslam
Department of Human Sciences, Loughborough University,
Loughborough, Leicestershire, UK

Introduction

Low back pain remains one of medicine's most enigmatic problems, particularly in its chronic form. The majority of people will experience back pain at some point in their lives [1], but only a minority will receive a positive diagnosis [2, 3]. Most will receive what are often perceived as

vague diagnoses, such as "mechanical back pain" or "non-specific low back pain". These apparently benign labels, however, hide a health problem that costs society more than cancer, coronary artery disease and AIDS combined [4]. The cost of this 'epidemic' [5] has spurred much debate amongst researchers, therapists and clinicians alike, with a common aim of defining effective treatment regimes. Bio-medical approaches have tended to be replaced by more 'holistic' bio-psychosocial approaches [6] and, indeed, some have argued that it is best not to view chronic back pain as a 'medical' problem at all [7]. There is certainly no one therapeutic 'silver bullet'; consequently, back pain is increasingly being treated as a multi-disciplinary problem.

Rehabilitation programmes tend to be built around the proven effectiveness of cognitive behavioural approaches [8–10] and exercise [11, 12], and they are typically delivered by multi-disciplinary teams. A review by Guzman et al. [13] showed the effectiveness of multi-disciplinary programmes, but a Cochrane Review [14] cautioned that many studies had methodological shortcomings. Some studies have also questioned the cost-effectiveness of such programmes [15]. Whilst the content of multi-disciplinary programmes can vary [16], the evidence base is sufficient for them to be considered the preferred choice for chronic non-specific back pain within recent clinical guidelines [17, 18].

One such multi-disciplinary programme is the 'Nottingham Back Team', which is a rehabilitation programme utilising a team of staff with nursing, physiotherapy, occupational therapy and clinical psychology skills. The team takes a cognitive behavioural approach based on a programme of education, goal setting, support and exercise. This paper describes an analysis of SF-36 data gathered within the programme immediately after its inception. The primary aim of this analysis was to evaluate the impact of the programme on participants' health 'profile' and the degree to which improvements were maintained at the 6-month follow-up. An additional objective was to identify any health differences between those completing the programme and those who either failed to attend or dropped-out after 2–3 weeks (which was relatively common). Following the approach described by Ferguson et al. [19], the Reliable Change Index (RCI) was used to gauge 'clinical significance' and to enable an analysis of individual scores in addition to group comparisons.

Materials and methods

Patients and methods

Patients referred for an assessment to the Back Team programme completed a questionnaire battery (including

the SF-36) prior to a physical examination and face-to-face consultation. The questionnaire battery was also administered at programme completion (at the end of the final session) and at the 6-month follow-up. Consent for use of the questionnaire data was obtained at the assessment visit.

The Nottingham Back Team was set up in 2000 in response to the Clinical Standards Advisory Group Report [20] and lengthy local waiting times for consultation at the spinal unit [21]. The programme is of fixed duration, consisting of seven 3-h sessions (morning or afternoon) on consecutive weeks. The sessions are held at local leisure centres within the community rather than being hospital-based. Each session has four key components:

- group education, covering a different topic each week (e.g. types of medication)
- group exercise—this varies from week to week, but with a core set of stretches/exercises on which participants track their progress (individuals are also provided with a tailored exercise programme to perform at home)
- relaxation training—after the exercise session, participants are taught a relaxation technique each week
- key worker—each participant has a designated key worker for the duration of the programme. On an individual basis, participants discuss their progress during the previous week and define goals for the following week.

In addition to the group sessions, participants have access (if required) to specialist support in relation to medication issues, cognitive behavioural therapy, vocational guidance and psychological assessment. Whilst the programme is of fixed duration, patients are offered an 'open appointment' for 1 year should they need additional support or advice following the programme.

The Short Form (SF)-36 Health Survey [22] is one of the most widely used health and quality of life measures. It's 36 items measure eight multi-item variables: physical functioning (PF); role limitation due to physical problems (RP); bodily pain (BP); general health (GH); vitality (VT); social functioning (SF); role limitation due to emotional problems (RE); mental health (MH). Version 2 of the SF-36 has been designed to facilitate norm-based scoring for all scales [23]. Raw scores are first transformed into 0–100 scores before being further transformed into T-scores where the mean is fixed as 50 (the population norm) and the standard deviation is 10.

Statistical analysis

The SF-36 data were analysed in terms of (1) statistical significance (change within and between groups); (2) with reference to the SF-36 scoring manual [23], which provides

tables of sample sizes needed to detect particular differences over time; (3) in relation to UK norms (T-scores); (4) utilising the RCI scores to assess 'clinical significance' (meaningful individual change).

The majority of reported studies using SF-36 describe analyses with parametric statistics [24]. It has been argued that non-parametric methods should be used [25] as the distributions tend to be skewed, but in reality, the use of parametric tests is unlikely to produce misleading results provided the samples are not small [24, 26]. When distributions are skewed, however, non-parametric methods can provide greater statistical power than parametric equivalents [27–29]. For this reason, group differences were analysed using Wilcoxon signed ranks (changes over time) and Mann–Whitney U (between groups differences) tests.

A number of methods are available to assess statistically significant individual change [30] and clinical significance/meaningful change [31]. Ferguson et al. [19] describe the use of RCI with SF-36 data, and the approach they used has been adopted within this study, but using relevant UK norm data [32, 33]. Clinical significance is taken as a change exceeding the RCI value that also results in the final score falling within the 'normal range' [19]. The cut-off defining 'normal range' can be taken as one, one and a half or two standard deviations from the norm [19]. The choice of cut-off will depend on the situation in question, but much of the literature on the RCI uses two standard deviations for studies in mental health [34].

Results

A total of 261 patients provided usable SF-36 questionnaires. Of these patients, 104 went on to complete the programme, whereas 112 failed to attend at all or failed to complete the programme. The remainder of the patients were unsuitable for the programme; this unsuitability may have been due to the nature of their condition (requiring alternative treatments), their level of disability (precluding participation in exercise elements) or 'logistical' problems, such as an inability to organise child care. Of the 104 patients completing the programme, 73 provided follow-up data at 6 months.

The 'health profile' of those considered suitable for the programme is illustrated in Fig. 1. The relationship to UK norms is illustrated as this figure uses data transformed to T-scores with a mean of 50 (equivalent to UK norm) and a standard deviation of 10.

The male:female ratio at assessment was 1:1. The age distribution of the patient sample was similar between genders, with female participants having a mean age of 43.8 years (SD 13.83) and male participants having a mean age of 44.5 years (SD 13.92). The even gender split was maintained at programme completion, but not at follow-up. At follow-up, the sample was 57% female.

Comparison of patients completing the programme with those who declined/dropped out

When comparing completers and non-completers, significant differences were found in each of the eight SF-36 scales (Table 1).

Changes in those completing the programme

For those completing the programme, significant improvements were evident on each scale (all $P < 0.000$). The SF-36 manual provides tables of sample sizes required to identify given changes in scores. Based on these tables, the sample sizes in our study were clearly large enough, given the size of changes observed. The changes in Physical Functioning and Role Physical scores exceeded RCI values. The changes are summarised in Table 2.

Figure 2 illustrates the change in health profile as a result of the programme; again, T-values are used to facilitate comparison with UK norms.

For those attending the 6-month follow-up, no significant group differences were evident, for any scale, between scores at programme completion and scores at follow-up.

Gender differences were evident at assessment, programme completion and follow-up. At assessment, there was a difference in Role Physical ($P < 0.05$); at programme completion, there were differences in Bodily Pain ($P < 0.05$) and Social Functioning ($P < 0.05$); at follow-up, there were differences in Bodily Pain ($P < 0.01$), Role Physical ($P < 0.05$), Social Functioning ($P < 0.05$) and Role Emotional ($P < 0.05$). In each case, the female scores exceeded the male scores.

Individual level

Table 3 shows the proportion of participants who completed the programme who showed 'clinically significant' improvements (i.e. exceeding RCI values) in each of the SF-36 scales.

Table 4 provides an overview of clinically significant changes for those participants attending the 6-month follow-up appointment in terms of three time periods—assessment to programme completion, programme completion to the 6-month follow-up and assessment to the 6-month follow-up. For the period programme completion to 6-month follow-up, there were clinically significant declines as well as improvements. There were no clinically significant declines for the other two time periods.

Fig. 1 The Short Form (SF)-36 health profile at the assessment of all patients suitable for the programme
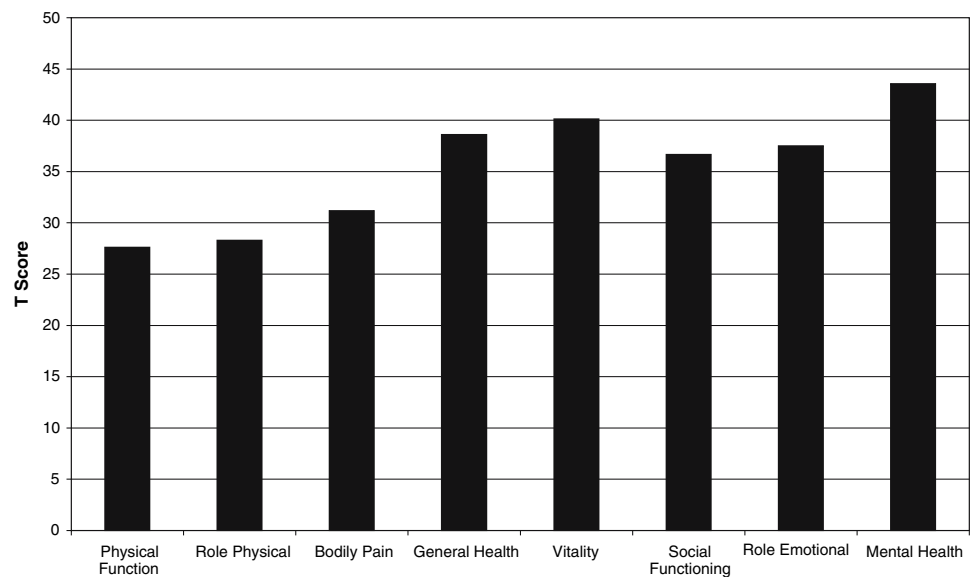
Table 1 Comparison of completers and non-completers (T-scores)

SF-36 scales	Non-completers	Completers
Physical functioning	24 (15)	31 (12)**
Role-physical	25 (15)	32 (14)**
Bodily pain	29 (10)	34 (10)**
General health	36 (10)	41 (10)**
Vitality	38 (11)	42 (11)*
Social functioning	35 (13)	39 (11)*
Role emotional	34 (17)	42 (14)**
Mental health	42 (13)	45 (10)*

*P < 0.05, **P < 0.01

Table 2 Improvement in SF-36 scores from assessment to programme completion (T-scores)

SF-36 scales	Assessment	Completion	Change (%)
Physical functioning	31 (12)	41 (12)	+32*
Role-physical	32 (14)	43 (12)	+33*
Bodily pain	34 (10)	43 (11)	+27*
General health	41 (10)	47 (10)	+14*
Vitality	42 (11)	49 (11)	+18*
Social functioning	39 (11)	47 (11)	+20*
Role emotional	42 (14)	48 (10)	+16*
Mental health	45 (10)	50 (11)	+11*

*P < 0.000

Table 5 describes the extent of clinically significant, concurrent improvement in the SF-36 scales. It describes the proportion of those participants showing a clinically significant improvement between assessment and the 6-month follow-up in each of the SF-36 scales who also

showed a clinically significant improvement in each of the other seven scales over the same period.

Table 6 describes the mean scores at assessment for each of the eight SF-36 scales for the whole sample and three important sub-groups. Those patients who showed at 6-month follow-up were found to have:

- a significant reduction in disability (physical functioning)
- a significant improvement in perceived functional limitation (role physical)
- a significant reduction in pain and related limitations (bodily pain).

This table illustrates differences between the whole sample (all those considered suitable for the programme) and those that showed clinically significant improvement between assessment and follow-up in physical functioning (CS-PF), role physical (CS-RP) and bodily pain (CS-BP). None of the differences were statistically significant however.

In terms of those who attended follow-up, only one participant showed no clinically significant change on any scale over any of the time periods involved in the study.

Of the follow-up group, 93% of participants showed clinically significant improvement between assessment and follow-up on at least one scale.

Discussion

The aim of this study was to evaluate both the impact of a rehabilitation programme on the health profile of patients who had enrolled in it and the degree to which any improvements were maintained over time and to explore

Fig. 2 Improvement in the health profile of the patients from assessment (*black bars*) to programme completion (*grey bars*) (T-scores)
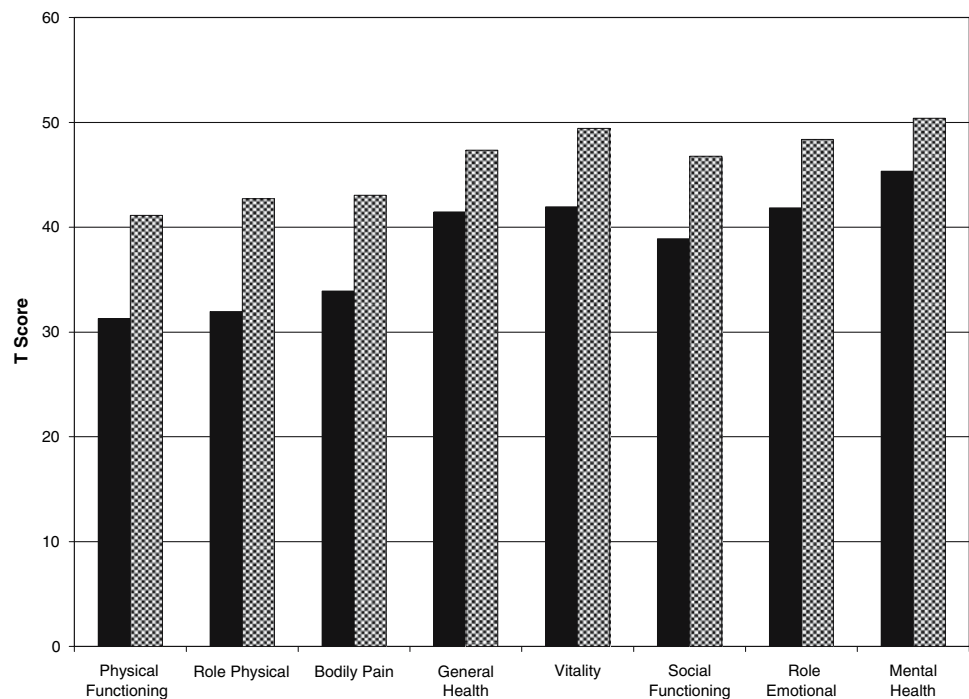
Table 3 Proportion of participants showing clinically significant improvement from assessment to programme completion, by gender

SF-36 scales	Males (%)	Females (%)	All (%)
Physical functioning	50	59	55
Role physical	56	55	55
Bodily pain	37	43	40
General health	23	22	23
Energy/vitality	33	31	32
Social functioning	44	41	43
Role emotional	42	39	41
Mental health	27	29	28

differences between those patients who completed the course and those that chose not to. The study utilised both individual and group scores in its analysis of an SF-36 data set. Whilst this is relatively unusual, both approaches are required in interpreting clinical significance in quality-of-life data [35]. Just as taking different statistical approaches to data analysis can alter the emphasis of results [36], using individual assessment alongside more familiar group comparisons allows a different perspective to be taken on the data. Individual assessment criteria also allow for the definition of sub-groups, which may prove useful in clinical practice by facilitating more tailored treatment regimes [37].

A number of methods are available for evaluating significant change at an individual level. Hays et al. [30] describe the use of the standard error of measurement and

the standard error of prediction alongside the RCI. These researchers emphasise the point that whichever measure is used, changes required to be significant at an individual level are much greater than those required at a group level. In fact, if samples are large enough, quite small group-level changes may be statistically significant, but essentially meaningless in clinical terms. Different techniques will provide different results [31]; however, there does appear to be a reasonable comparability between approaches [38, 39] that is supportive of the construct as a whole. Although no single approach is likely to be uniformly accepted [40] and 'blindly' following a single threshold value could be misleading [41], the emerging evidence is supportive of the view that 'clinical significance' is an important element in the evaluation of treatment effectiveness.

Using Reliable Change as an evaluation criterion was first suggested over 20 years ago [42], and the technique was described in detail by Jacobson and Truax in 1991 [43]. Since then it has been increasingly used in fields such as psychotherapy, but it is much less common in medical outcomes research [19]. The concept of defining a minimum criterion for 'clinically significant' change is not new in disability or rehabilitation research however. For example, it is widely accepted that a change of less that 2.5 points on the Roland–Morris Disability Questionnaire should not be seen as 'clinically significant' [44, 45].

Ferguson et al. [19] describe the application of RCI to SF-36 outcomes, and the approach they defined has been taken in this study, but with RCI criteria calculated using relevant UK data from the Health Services Research Unit

Table 4 Proportion of participants showing clinically significant changes

SF-36 scales	Improvement from assessment to programme completion (%)	Improvement from programme completion to follow-up (%)	Decline from programme completion to follow-up (%)	Overall improvement from assessment to follow-up (%)
Physical functioning	66	8	15	65
Role physical	63	15	17	70
Bodily pain	46	6	13	49
General health	25	3	4	21
Energy/vitality	34	8	7	38
Social functioning	44	10	14	41
Role emotional	45	15	14	42
Mental health	32	6	8	24

Table 5 Proportion of those patients showing a clinically significant (CS) improvement in each scale who also showed a concurrent improvement in other scales

For those showing CS improvement in:	Proportion (%) also showing clinically significant improvement in:							
	PF	RP	BP	GH	VT	SF	RE	MH
Physical functioning (PF)	–	83	65	28	46	52	48	33
Role physical (RP)	76	–	64	26	44	48	52	50
Bodily pain (BP)	86	91	–	23	51	63	49	29
General health (GH)	80	80	47	–	60	67	53	47
Vitality (VT)	78	85	67	37	–	59	63	52
Social functioning (SF)	83	79	76	38	65	–	55	45
Role emotion (RE)	70	83	57	27	57	50	–	40
Mental health (MH)	88	71	59	47	82	76	71	–

Table 6 Comparison of mean scores at assessment (T-scores)

SF-36 scales	All suitable	CS-PF	CS-RP	CS-BP
Physical functioning	30	27	28	31
Role-physical	30	29	27	31
Bodily pain	32	31	31	31
General health	42	43	43	44
Vitality	42	42	42	43
Social functioning	39	37	38	39
Role emotional	41	39	40	41
Mental health	45	44	45	46

at the University of Oxford [32, 33]. Whilst any particular defined change value can be said to be somewhat arbitrary, the RCI is a conservative measure [35]. The SF-36 is itself a robust tool [22], and its scales are associated with low back pain outcome measures [46]. Hence, the combination of the two should prove effective even if it may tend to underestimate the impact of treatment.

Both the assessment and post-programme groups showed an even gender split. Previous studies have shown differences in pain perception and associated behaviours between genders [47, 48] which have been explained in terms of psychological, physiological and socio-cultural factors. Similar differences have been described for treatment outcomes [49, 50]. In this study, gender differences were identified in four SF-36 scales—Role Physical, Bodily Pain, Social Functioning and Role Emotional. In each case, male participants scored lower than their female counterparts. This is the opposite of most previous research where females tend to report higher pain levels [51]. The reasons for this anomaly are unclear, but it appears that female patients respond well to this programme, with proportionately more females showing clinically significant improvement in physical functioning and bodily pain.

Predictably, given the health profile of those assessed and deemed suitable for the Back Team programme, the lowest scores were found for Physical Functioning, Role Physical and Bodily Pain, with each of these scales falling more than one and a half standard deviations from the norm. The only other scale falling further than one standard deviation from the norm was Social Functioning. The scale falling closest to the norm was Mental Health, which was within half a standard deviation, despite the fact that low back pain is often associated with psychological factors

[52–54]. Indeed, the mean mental health score at programme completion was very slightly above UK norm. An ongoing research project will be looking at these issues.

When we compared the scores of participants at assessment, with those participants who subsequently completed the programme and those who failed to complete the programme, we found statistically significant differences on all scales. The differences were relatively small however, compared to the size of change in response to the programme. The scale Role Physical was found to exceed the RCI value. It would appear that failure to engage with the programme may be due to a belief amongst potential participants that they are too disabled by their condition to undertake a programme of this type. This could reflect strong 'organic' pain beliefs [55], where the patient strongly believes 'hurt = harm' (a 'biomedical' perspective). Patients with strong organic pain beliefs tend to have an external locus of control and are more likely to expect the medical practitioner to 'cure' them rather than take responsibility for themselves [55]. The nature of pain beliefs within this population is being examined within further research with the programme.

Equally, failure to engage may reflect a lack of understanding of what the programme will entail, and some similar programmes have successfully utilised a pre-programme session to deal with this. Unfortunately, resource constraints have prevented the team from gathering data on the reasons for failure to engage with the programme. More recently, the team has changed its policy such that places are not routinely allocated to those assessed as suitable for the programme; instead, patients are asked to explicitly 'opt in' to a particular group. This new approach has led to fuller groups, improved completion rates and reduced waiting times, but it has not addressed the underlying issue of why some individuals fail to engage.

For those completing the programme, statistically significant improvements were found on all scales. Based on the SF-36 scoring manual, the sample size exceeded the 'numbers needed' for the magnitude of change. Only two scales exceeded RCI criteria—Physical Functioning and Role Physical. At a group level, this programme can be said to provide clinically significant improvements in disability for those completing the programme.

Whilst traditional statistical analysis can provide scientifically robust results, a comparison of two distributions provides very little information on the extent to which any individual improves or declines. By defining clinically significant change with RCI values, we obtained a different perspective on the data. Using this approach, it can be seen that of those completing the programme, the scales on which most of the participants improved were Physical Functioning and Role Physical,

with more than half the sample showing clinically significant improvement. After these scales, Bodily Pain, Social Functioning and Role Emotional showed a similar proportion of the sample improving at around 40%. General Health, Vitality and Mental Health scales showed lower proportions of improvement at 23, 32 and 28%, respectively.

An analysis of the data from those attending the 6-month follow-up provided an understanding of change across three important time frames:

- the duration of the programme—difference between the assessment and programme completion scores
- change subsequent to the programme—difference between the programme completion score and 6-month follow-up
- Overall, long-term change—the difference between the initial assessment scores and those at follow-up.

For six of the eight scales, more participants showed a decline between programme completion and follow-up than an improvement, the exceptions being for Vitality and Role Emotional. The largest proportional decline between programme completion and follow-up occurred in the scales which had shown the largest numbers improving from assessment to programme completion—Physical Functioning and Role Physical. These figures indicate that change can take place in a number of ways:

- initial improvement, maintained at follow-up ('early takers')
- initial improvement, lost at follow-up ('need support')
- no initial improvement, but improvement by follow-up ('slow burners')
- no improvement at any stage ('non-takers').

These results have important implications when attempting to assess the true impact of a programme as those showing improvement at programme completion may not maintain that improvement once the support of the programme is removed. Equally, for the 'slow burners', a rehabilitation programme may facilitate a major change in their life (the beginnings of a recovery) without showing a significant improvement during the course of the programme itself. When the cost benefits of this type of programme are being considered, consideration must be given to the degree to which gains during the programme may be lost following its completion and to the optimal nature of the follow-up and support, which may need to be better tailored to the needs of particular patient groups (particularly the 'need support' group).

Analysing individual change also illustrates that improvements are not seen uniformly across scales. In terms of 'long-term' change (assessment to follow-up), the scale on which the greatest number of participants

improved (70%) was Role Physical. This could be said to reflect the cognitive behavioural aspect of the course whereby participants learn that they can manage their condition and that it need not control or restrict their lives. Almost a quarter of those showing an improvement in Role Physical did not show a corresponding improvement in Physical Functioning however, suggesting that improvement in 'perceived' disability can be made without a real change in the limitation imposed by their condition. This result supports the view that disability is driven by psychological as well as physical factors [56, 57].

Whilst the figure of 70% is the highest score for an individual scale, this could be said to under-estimate the impact of the programme as 93% of those attending the 6-month follow-up showed an improvement in at least one scale (assessment to follow-up). For the overwhelming majority of the participants continuing the programme through to follow-up, the programme can be said to have had a genuine positive effect. Whilst some observed improvements at programme completion were subsequently lost to follow-up, only one participant failed to show a clinically significant change on any scale over any time period. Although some of the improvement may be due more to the social nature of the group programme than to its content, the benefits in quality of life are nevertheless important and 'real'. Previous work by the lead author [58] has shown the impact of social isolation amongst chronic back pain sufferers and the benefits perceived within a group setting.

Over a third of those showing clinically significant change in the scales Physical Functioning and Role Physical do not show a corresponding improvement in Bodily Pain scores. This suggests that disability can improve without a reduction in pain levels and supports previous research showing the lack of a uniform relationship between pain levels and disability [59].

The individual scores can also be used to allow subgroups to be defined. By splitting our patient sample into groups based on whether they had shown a clinically significant change on a particular variable, we sought clues for some predisposition to improvement. When we considered the mean scores at assessment for each SF-36 scale for the whole sample and for those who had shown clinically significant improvement in Physical Functioning, Role Physical and Bodily Pain, we found little difference, with the greatest differences being on the General Health scale. Even for this scale, however, there were no statistically significant differences between those who subsequently showed clinically significant improvement and those who did not. Beliefs in relation to 'general health' and 'perceived disability' may be significant in programme completion, however, and this area warrants further research.

Conclusions

This study illustrates the benefits to be gained by analysing individual change in addition to group changes when evaluating the impact of a rehabilitation programme. From both perspectives, the programme assessed in this paper had its greatest impact on Physical Functioning and Role Physical, which are the two SF-36 scales most closed related to disability. The improvement was not a simple linear change, however, as some participants showed a loss of improvement at follow-up whilst others showed an improvement at follow-up without one being evident over the duration of the programme itself.

Equally, the results of this study illustrate that an improvement in one scale is not necessarily associated with improvements on other scales: we found that an improvement in pain was associated with an improvement in disability for the vast majority of cases, but the reverse was not true. This study therefore supports the view that improving an individual's pain score is not a pre-requisite to addressing disability levels.

The programme had widespread impact, and the vast majority of participants at follow-up had shown clinically significant improvement in at least one scale between assessment and follow-up. Whilst this is not a controlled experimental study, the data suggests that the Back Team Programme represents an effective intervention for those with chronic low back pain.

Very few participants have been unaffected by the programme, so efforts to improve its effectiveness should concentrate on addressing decline following the programme and in ensuring improvement over a wider spectrum of health factors. Of greatest concern, however, is the proportion of those suitable for the programme who decline the opportunity. Further research should explore the factors affecting this decision and other differences between completers and non-completers with a goal of improving programme uptake.

References

1. Waddell, G. (2004). *The back pain revolution*. London: Churchill Livingstone.
2. Snook, S. (2004). Work-related low back pain: Secondary intervention. *Journal of Electromyography Kinesiology, 14*, 153–160.
3. Deyo, R., & Weinstein, J. (2001). Low back pain. *The New England Journal of Medicine, 344*, 363–370.
4. Thomsen, A., Sorenson, J., Sjorgen, P., & Eriksen, J. (2002). Chronic non-malignant pain patients and health economic consequences. *European Journal of Pain, 6*, 341–352.

5. Bruntland, G. (2000). Scientific group meeting on the burden of musculoskeletal disorders. World Health Organisation. http://www.who.int/director-general/speeches/2000/english/20000113_bone_joint.html. Accessed 20 August 2000.

6. Borkan, J., Van Tulder, M., Reis, S., Schoene, M., Croft, P., & Hermoni, D. (2002). Advances in the field of low back pain in primary care: A report from the fourth international forum. *Spine, 27*, 128–132.

7. Fordyce, W., & the International Association for the Study of Pain Task Force on Pain. (1995). *Back pain in the workplace: Management of disability in nonspecific conditions*. Seattle: IASP Press.

8. Linton, S., & Ryberg, M. (2001). A cognitive-behavioral group intervention as prevention for persistent neck and back pain in a non-patient population: A randomized controlled trial. *Pain, 90*, 83–90.

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

10. van Tulder, M., Ostelo, R., Vlaeyen, J., Linton, S., Morley, S., & Assendelft, W. (2000). Behavioral treatment for chronic low back pain: A systematic review within the framework of the Cochrane Back Review Group. *Spine, 25*, 2688–2699.

11. Abenhaim, L., Rossignol, M., Valat, J., Nordin, M., Avouac, B., & Blotman, F. (2000). The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. *Spine, 25*, 1S–33S.

12. Liddle, S., Baxter, G., & Gracey, J. (2004). Exercise and chronic low back pain: What works? *Pain, 107*, 176–190.

13. Guzman, J., Esmail, R., Karjalainen, K., Malmivaara, A., Irvin, E., & Bombardier, C. (2001). Multidisciplinary rehabilitation for chronic low back pain: Systematic review. *British Medical Journal, 322*, 1511–1516.

14. Karjalainen, K., Malmivaara, A., van Tulder, M., Roine, R., Jauhiainen, M., Hurri, H., & Koes, B. (2001). Multidisciplinary biopsychosocial rehabilitation for subacute low back pain in working-age adults: A systematic review within the framework of the Cochrane Collaboration Back Review Group. *Spine, 26*, 262–269.

15. Turk, D. (2002). Chronic non-malignant pain patients and health economic consequences. *European Journal of Pain, 6*, 353–355.

16. Airaksinen, O., Brox, J.-I., Cedraschi, C., Hildebrandt, J., Klaber-Moffett, J., Kovacs, F., Mannion, A., Reis, S., Staal, J., Ursin, H., & Zanoli, G. (2005). Multi-disciplinary interventions. In Cost Action B13, *European guidelines for the management of chronic nonspecific low back pain*. Brussels: European Commission Research Inspectorate General.

17. Balague, F., Mannion, A., Pellise F., & Cedraschi, C. (2007). Clinical update: Low back pain. *The Lancet, 369*, 726–728.

18. Chou, R., Qaseem, A., Snow, V., Casey, D., Cross, J., Shekelle, P., & Owens, D. (2007). Diagnosis and treatment of low back pain: A joint clinical practice guideline from the American college of physicians and the American pain society. *Annals of Internal Medicine, 147*, 478–491.

19. Ferguson, R., Robinson, A., & Splaine, M. (2002). Use of the Reliable Change Index to evaluate clinical significance in SF-36 outcomes. *Quality of Life Research, 11*, 509–516.

20. Clinical Standards Advisory Group Committee on Back Pain. (1994). *Back pain: Report of a CSAG committee on back pain*. London: HMSO.

21. The Improvement Network—East Midlands. (n.d.). The Nottingham back team. Retrieved Aug 27, 2005 from http://www.tin.nhs.uk/welcome/good-news-stories/back-pain-team

22. Ware, J. (2000). SF-36 health survey update. *Spine, 25*, 3130–3139.

23. Ware, J., Kosinki, M., Bjorner, J., Turner-Bowker, D., Gandek, B, & Maruish, M. (2007). *User's manual for the SF-36v2 health survey* (2nd ed.). Lincoln, RI: QualityMetric Incorporated.

24. Williams, P., & Nerenz, D. (1995). When should nonparametric statistics be used to analyses f-36 scores? *AHSR FHSR Annual Meeting Abstract Book, 12*, 152–153.

25. Velanovich, V. (2007). Behaviour and analysis of sf-36 short-form health survey data for surgical quality-of-life research. *Archives of Surgery, 142*, 473–478.

26. Walters, S., & Campbell, M. (2004). The use of bootstrap methods for analyzing health-related quality of life outcomes (particularly the SF-36). *Health and Quality of Life Outcomes, 2*, 70.

27. Walters, S. (2004). Sample size and power estimation for studies with health related quality of life outcomes: A comparison of four methods using SF-36. *Health and Quality of Life Outcomes, 2*, 26.

28. Vickers, A. (2005). Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Medical Research Methodology, 5*, 35.

29. Bridge, P., & Sawilowsky, S. (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: Comparative Power of the *t*-test and Wilcoxon rank–sum test in small samples applied research. *Journal of Clinical Epidemiology, 52*, 229–235.

30. Hays, R., Brodsky, M., Johnston, M., Spritzer, K., & Hui, K. (2005). Evaluating the statistical significance of health-related quality-of-life change in individual patients. *Evaluation & Health Professions, 28*, 160–171.

31. Bauer, S, Lambert, M., & Nielsen, S. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment, 82*, 60–70.

32. Jenkinson, C., Stewart-Brown, S., Peterson, S., & Paice, C. (1999). Assessment of the SF-36 version 2 in the United Kingdom. *Journal of Epidemiology and Community Health, 53*, 46–50.

33. Jenkinson, C., Stewart-Brown, S., & Peterson, S. (2005). *Assessment and evaluation of the SF36 Version II*. Health Services Research Unit, University of Oxford. http://www.hsru.ox.ac.uk/sf36v2.htm. Accessed 2 July 2005.

34. Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health, 1*, 70–72.

35. Cella, D., Bullinger, M., Scott, C., & Barofsky, I. (2002). Group vs. individual approaches to understanding the clinical significance of differences or changes in quality of life. *Mayo Clinic Proceedings, 77*, 384–392.

36. Hoogendoorn, W., Bongers, P., de Vet, H., Twisk, J., van Mechelen, W., & Bouter, L. (2002). Comparison of two different approaches for the analysis of data from a prospective cohort study: An application to work related risk factors for low back pain. *Occupational and Environmental Medicine, 59*, 459–465.

37. Vollenbroek-Hutten, M., Hermens, H., Wever, D., Gorter, M., Rinket, J., & Ijzerman, M. (2004). Differences in outcome of a multi-disciplinary treatment between subgroups of chronic low back pain patients defined using two multiaxial assessment instruments: The multidimensional pain inventory and lumbar dynamometry. *Clinical Rehabilitation, 18*, 566–579.

38. Atkins, D., Bedics, J., McGlinchey, J., & Beauchaine, T. (2005). Assessing clinical significance: Does it matter which method we use? *Journal of Consulting and Clinical Psychology, 73*, 982–989.

39. Sloan, J., Cella, D., & Hays, R. (2005). Clinical significance of patient-reported questionnaire data: Another step towards consensus. *Journal of Clinical Epidemiology, 58*, 1217–1219.

40. Sloan, J. (2003). Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. *Drug Information Journal, 37*, 23–31.

41. Hays, R., & Woolley, J. (2000). The concept of clinically meaningful difference in health-related quality-of-life research—how meaningful is it? *Pharmacoeconomics, 18*, 419–423.

42. Jacobson, N., Follette, W., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behaviour Therapy, 15*, 336–352.

43. Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.

44. Bombardier, C., Hayden, J., & Beaton, D. (2001). Minimal clinically important difference. Low back pain: Outcome measures. *The Journal of Rheumatology, 28*, 431–438.

45. UK Beam Trial Team. (2004). United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: Effectiveness of physical treatments for back pain in primary care. *British Medical Journal, 329*, 1377.

46. Gatchel, R., Mayer, T., Dersh, J., Robinson, R., & Polatin, P. (1999). The association of the SF-36 health status survey with 1-year socioeconomic outcomes in a chronically disabled spinal disorder population. *Spine, 24*, 2162–2170.

47. Keogh, E., & Herdenfeldt, M. (2002). Gender, coping and the perception of pain. *Pain, 97*, 195–201.

48. Dixon, K., Thorn, B., & Ward, L. (2004). An evaluation of sex differences in psychological and physiological responses to experimentally-induced pain: A path analytic description. *Pain, 112*, 188–196.

49. Keogh, E., McCracken, L., & Eclestone, C. (2005). Do men and women differ in their response to interdisciplinary chronic pain management? *Pain, 114*, 37–46.

50. McGeary, D., Mayer, T., Gatchel, R., Anagnostis, C., & Proctor, T. (2003). Gender-related differences in treatment outcomes for patients with musculoskeletal disorders. *The Spine Journal, 3*, 197–203.

51. Berkeley, K. (1997). Sex differences in pain. *Behavioural and Brain Sciences, 20*, 371–380.

52. Ciaramella, A., Grosso, S., Poli, P., Gioia, A., Inghirami, S., & Massimetti, G. (2004). When pain is not fully explained by organic lesion: A psychiatric perspective on chronic pain patients. *European Journal of Pain, 8*, 13–22.

53. Crook, J., Milner, R., Schultz, Z., & Stringer, B. (2002). Determinants of occupational disability following a low back injury: A critical review of the literature. *Journal of Occupational Rehabilitation, 12*, 277–295.

54. Linton, S. (2000). A review of psychological risk factors in back and neck pain. *Spine, 25*, 1148–1156.

55. Edwards, L., Pearce, S., Turner-Stokes, L., & Jones, A. (1994). The pain beliefs questionnaire: An investigation of beliefs in the causes and consequences of pain. *Pain, 51*, 267–272.

56. Crombez, G., Vlaeyen, J., Heuts, P., & Lysens, R. (1999). Pain-related fear is more disabling than pain itself: Evidence on the role of pain-related fear in chronic back pain disability. *Pain, 80*, 329–339.

57. Schultz, I., Crook, J., Meloche, G., Berkowitz, J., Milner, R., & Zuberbier, O. (2004). Psychosocial factors predictive of occupational low back disability: Towards development of a return-to-work model. *Pain, 107*, 77–85.

58. Baird, A. (2004). *The impact of pain beliefs and fear avoidant behaviours on perceived pain and physical capabilities—a preliminary study*. Presented at the 37th Pain Society Annual Scientific Meeting, Manchester, UK.

59. Turner, J., Franklin, G., Heagerty, P., Wu, R., Egan, K., Fulton-Kehoe, D., Gluck, J., & Wickizer, T. (2004). The association between pain and disability. *Pain, 112*, 307–314.